(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 February 2002 (14.02.2002)

PCT

(10) International Publication Number
WO 02/12335 A2

(51) International Patent Classification[7]: C07K 14/485

(21) International Application Number: PCT/US01/25282

(22) International Filing Date: 10 August 2001 (10.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/636,474 10 August 2000 (10.08.2000) US

(71) Applicant *(for all designated States except US)*: **PHARMACIA CORPORATION** [US/US]; 800 North Lindbergh Boulevard, Mail Zone O4E, St. Louis, MO 63167 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **DOSHI, Parul, D.** [IN/US]; 16763 Hickory Crest Dr., Wildwood, MO 63011 (US). **MEIER, Martin, F.** [US/US]; 15801 H. Timbervalley Rd., Chesterfield, MO 63017 (US). **WOODS, Cynthia, L.** [US/US]; 425 Wyatt Dr., St. Peters, MO 63376 (US).

(74) Agent: **BAUER, S., Christopher**; Pharmacia Corporation, Corporate Patent Department, 800 North Lindbergh Blvd., Mail Zone O4E, St. Louis, MO 63167 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/12335 A2

(54) Title: A NEW EGF MOTIF REPEAT PROTEIN OBTAINED FROM A cDNA LIBRARY FROM HS-5 STROMAL CELL LINE

(57) Abstract: The present invention relates to a new EGF motif repeat protein obtained from a cDNA library from HS-5 stromal cell line, nucleic acid molecules that encode the EGF motif repeat protein or a fragment thereof and nucleic acid molecules that are useful as probes or primers for detecting or amplifying the EGF motif repeat protein, respectively. The invention further relates to uses of the EGF motif repeat protein. The invention also relates to applications of the factor or fragment such as forming antibodies capable of binding the factor or fragments thereof.

A NEW EGF MOTIF REPEAT PROTEIN OBTAINED FROM A cDNA LIBRARY FROM HS-5 STROMAL CELL LINE

The present application is a continuation-in-part of United States application Serial No. 09/636,477, filed August 10, 2000, which is incorporated by reference in entirety as if written herein.

FIELD OF THE INVENTION

The present invention relates to a new EGF motif repeat protein obtained from a cDNA library from the stromal cell line, HS-5, nucleic acid molecules that encode the EGF motif repeat protein or a fragment thereof, and nucleic acid molecules that are useful as probes or primers for detecting or amplifying the factor, respectively. The invention further relates to uses of the EGF motif repeat protein. The invention also relates to applications such as forming antibodies capable of binding the factor or fragments thereof.

BACKGROUND

Epidermal growth factor (EGF) is a soluble peptide of 53 amino acids with a distinctive motif of six cysteines. The molecule is tightly folded into a structure with a distinctive amino-terminal triple-stranded β-sheet, which is stabilized by disulfide bonding of the cysteine residues. The hallmark of the EGF motif is the highly conserved spacing of six cysteine residues. This EFG motif is found in many different proteins of diverse function. Soluble EGF is derived from the proteolytic cleavage of a much larger precursor on the cell surface. Three other growth factors are extremely homologous to EGF. For transforming growth factor-α (TGFα), vaccinia virus growth factor (VVGF), and amphiregulin, the functional significance of their homology is clear, since all three factors bind to the EGF receptor. This EGF motif has been found repeated in many different proteins of diverse functions. Like EGF, both TGF-α and VVGF are derived from a transmembrane precursor.

The finding that TGF-α surface-bound EGF precursor domains can mediate intercellular communication has had important implications for the role of EGF motifs in the function of other cell surface molecules as well. EGF repeats are found in a number of different cell surface molecules that play roles in calcium binding, cell-cell interactions, and cell-fate decisions (see Davis, G., *The New Biologist* 2:410-419, 1990 for review). This diverse group includes: lipoprotein receptors (Yamamoto et al. *Cell* 39:27-38, 1984); thrombomodulin (Suzuki, K. et al., *EMBO* 6:1891-1897, 1987); selectins such as endothial leukocyte adhesion molecule-1 (ELAM-1) (Becvilacqua et al. *Science* 243:1160-1165, 1989); extracellular matrix proteins (see Engel, J., *FEBS Letters* 251:1-7, 1989 for review) including tenascins (Chiquet-Ehrisman R., *Experientia* 51:853-862, 1995), Fibulin (Pan, T. et al. *J. Cell Bio.* 123:1269-1277, 1993), lamain (Sasaki M. et al., *PNAS* 84:935-939, 1988) nidogen (Mann, K., *EMBO* 8:65-72, 1989) and thrombospondin (Majack, R. A., et al., *PNAS* 83:9050-9054, 1986); soluble secreted proteins, such as coagulation factors VII, IX, X, and protein C (see Furie and Furie, *Cell* 53:505-518, 1988 for review), and tissue-type plasminogen activator (t-PA) (Pennica, D. et al., *Nature* 301;214-221, 1983); Slit (Rothberg, J. M. et al., *Genes & Development* 4:2169-2187, 1990) and three mammalian Slit homologues (Itoh, A. et al., *Molecular Brain Research* 62:175-186, 1998; Connolly, T. and Rajuput, WO99/23219, 1999); the seven-transmembrane-region leukocyte cell-surface family (EGF-TM7) of molecules (see McKnight, A. J. and Gordon, S., *Immunology Today* 6:283-287, 1996 for review); and the Notch family of homeotic genes that control the developmental fate of cells with binary potential.

In *Drosophila*, Notch (Kidd, S. et al., *Mol. Cell Biol.* 6:3094-3108, 1986) and its ligands Delta (Vassin H., *EMBO* 6:3431-3440, 1987) and Serrate (Fleming R. J. et al., *Genes & Development* 4:1188-2201) are involved in neurological development. The products of two genes in *C. elegans*, *lin-12* and *glp-1* are closely related in structure to Notch. At present four Notch homologs have been identified in vertebrates Notch-1, -2, -3, and –4 (Ellison L. W. et al., *Cell* 66:649-661, 1991; Weinmaster, G. et al., *Development* 116:931-941, 1992; Lardelli, M. et al., *Mech Dev.* 46:123-126, 1994; Uyttendaele H. et al., *Development* 122:2251-2259, 1996). Also four Notch ligands referred to as Jagged and Delta-like (dlk) have been cloned

from vertebrates including Jagged-2 from humans (Luo, B. et al., *Mol Cell Biol*, 17:6057-6067, 1997). The Notch superfamily includes the human homolog of *Drosophila* Notch–1, TAN-1 (Ellisen, L. W., *Cell* 66:649-661, 1991), and its ligands Jagged-2 (Lou, B. et al., *Mol Cell Biol*, 17:6057-6067, 1997) and delta-like (dlk) (WO 94/13701, June 23, 1994).

In addition to its role in mediating cell fate decisions by multipotent precursors in a number of different systems including central and peripheral nervous system development, the Notch superfamily has been shown to mediate cell fate determination in hematopoiesis (see Milner L. A. and Bigas, A., *Blood* 93:2431-2448, 1999 for review). Hematopoiesis is a continuous development process in which pluripotent stem cells and their progeny make cell fate decisions, producing mature blood cells of the various lineages. The consent generation of appropriate numbers and types of mature cells, as well as the maintenance of pluripotent progenitors, requires a complex regulatory network (; Metcalf, D., *Stem Cells* 16(5) 314-321, 1998 (suppl 1); Metcalf, D., *Blood* 92:345-352, 1998). Since the initial demonstration that the human Notch homolog is expressed in $CD34^+$ hematopoietic precursors (Milner L. A. et al., *Blood* 83:2057-2062, 1994), considerable evidence has emerged to support a conserved role for the Notch superfamily in the mediation of cell fate decisions and self-renewal of progenitors during hematopoiesis. Jagged-1 has been shown to promote the proliferation of hematopoietic precursor populations (Varnum-Finney B. et al, *Blood* 11:4084-4091, 1998).

Evidence is accumulating for many tumors that chromosomal translocations play a direct role in neoplastic transformation and tumor progression. Genes located at or near the breakpoints of several common translocations in various forms of leukemia and lymphoma have been molecularly cloned, and it is apparent that in some cases these genes are proto-ongogenes related to previously identified genes in transforming retrovirus. In human acute lymphoblastic T-cell leukemia, a chromosomal translocation damages the *Notch-1* gene (Ellisen, L. W. et al., *Cell* 66:649-661, 1991; Capobianco, A. J. et al., *Molecular and Cellular Biology* 17:6265-6273, 1997).

A number of additional genes have been cloned from a fetal liver-spleen cDNA library encoding proteins with EGF motif repeats (WO 99/27096 and WO

00/17357) and a human brain cDNA library (Nakayama M. et al., *Genomics* 51:27-34, 1998).

Several approaches have been used to identify and analyze differentially expressed genes. These approaches include differential cDNA library screening, subtractive libraries and PCR-based differential display, single cell amplification, and PCR Select (Diatchenko *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 93*:6025-6030 (1996); Maser and Calvet, *Semin. Nephrol. 15*:29-42 (1995); Watson and Marguiles, *Dev. Neurosci. 15*:77-86 (1993); Miles and Wallace, *Behring. Inst. Mitt. 88*:133-141 (1991)).

Subtractive libraries are a classic approach to the isolation of novel proteins. The subtractive hybridization approach has been reviewed by Swendeman and La Quaglia, *Semin. Pedriatr. Surg. 5*:149-154 (1996); Ermolaeva and Sverdlov, *Genet. Anal. 13*:49-58 (1996); Hara *et al.*, *Anal. Biochem. 214*:58-64 (1993). Subtractive hybridization has been used to identify cDNA clones that are predominantly expressed in the immune system (Ericsson *et al.*, *Immunol. Rev. 100*:261-277 (1987)). With respect to stem cell growth factors, however, subtractive hybridization has been reported to be cumbersome.

The differential display approach has been reviewed by Wang and Feuerstein, *Cardiovasc. Res. 35*:414-421 (1997); Wang *et al.*, *Trends Pharmacol. Sci. 17*:276-279 (1996); Livesey and Hunt, *Trends Neurosci. 19*:84-88 (1996); Sunday, *Am. J. Physiol. 269*:L273-L284 (1995); and Liang and Pardee, *Curr. Opin. Immunol. 7*:274-280 (1995); Mou *et al.*, *Biochem. Biophys. Res. Commun. 199*:564-569 (1994). It has been reported that cytokine-inducible genes have been isolated from hematopoietic cells by the differential display approach (Zhu *et al.*, *Methods Mol. Biol. 85*:153-161 (1997); Rosok *et al.*, *Bio/Techniques 21*:114-121 (1996)).

PCR-Select (suppression subtractive hybridization (SSH)) combines normalization and subtraction in a single procedure (Diatchenko *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 93*:6025-6030 (1996); Gurskaya *et al.*, *Anal. Biochem. 240*:90-97 (1996); PCR-Select cDNA Subtraction Kit, *CLONTECHniques X(4)*:2-5 (1995)). The PCR-Select approach employs suppression PCR (U.S. Patent No. 5,565,340). A kit for PCR-Select is available from Clontech (Palo Alto, CA).

SUMMARY OF THE INVENTION

Based on a study of the HS-5 cell line, a library of 294 substantially purified nucleic acid sequences was generated. These sequences are derived from the genes produced by HS-5 and are useful as probes in detecting the entire gene and as primers for amplifying the gene sequence. The present invention also relates to a substantially-purified nucleic acid molecule that encodes an EGF motif repeat protein, a variant thereof, or fragment thereof and which comprises at least one nucleic acid sequence selected from the group consisting of SEQ ID NO:1 through SEQ ID NO:297, substantial fragments thereof, substantial homologues thereof, or substantial complements thereof.

The present invention provides a substantially-purified nucleic acid molecule that encodes an EGF motif repeat protein, or variant thereof, or fragment thereof, having a nucleic acid sequence that substantially hybridizes with at least one nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, and complements thereof, is also provided by the present invention.

The present invention further relates to a substantially-purified EGF motif repeat protein, or fragment thereof, encoded by a nucleic acid molecule which substantially hybridizes to a second nucleic acid molecule, where the second nucleic acid molecule consists essentially of a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, and complements thereof.

The present invention also provides a method for determining a level or pattern of an EGF motif repeat protein in a cell comprising: (A) incubating, under conditions permitting nucleic acid hybridization, a marker nucleic acid molecule, the marker nucleic acid molecule having a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or complements thereof, with a complementary nucleic acid molecule obtained from the cell, wherein nucleic acid hybridization between the marker nucleic acid molecule, and the complementary nucleic acid molecule obtained from the cell permits the detection of the EGF motif repeat protein; (B) permitting hybridization between the marker nucleic acid molecule and the complementary nucleic acid

molecule obtained from the cell; and (C) detecting the level or pattern of the complementary nucleic acid, wherein the detection of the complementary nucleic acid is predictive of the level or pattern of the EGF motif repeat protein.

The present invention also provides a method for determining a mutation in a cell whose presence is predictive of a mutation affecting the level or pattern of an EGF motif repeat protein comprising the steps: (A) incubating, under conditions permitting nucleic acid hybridization, a marker nucleic acid molecule, the marker nucleic acid molecule comprising a nucleic acid molecule that is linked to a gene, the gene specifically hybridizes to a nucleic acid molecule having a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or complements thereof, and a complementary nucleic acid molecule obtained from the cell, wherein nucleic acid hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the cell permits the detection of a polymorphism whose presence is predictive of a mutation affecting the level or pattern of the EGF motif repeat protein in the cell; (B) permitting hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the cell; and (C) detecting the presence of the polymorphism, wherein the detection of the polymorphism is predictive of the mutation.

The present invention also provides a method of producing a cell containing an overexpressed EGF motif repeat protein comprising: (A) transforming the cell with a functional nucleic acid molecule, wherein the functional nucleic acid molecule comprises a promoter region, wherein the promoter region is linked to a structural region, wherein the structural region comprises a nucleic acid molecule having a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or complements thereof; wherein the structural region is linked to a 3' non-translated sequence that functions in the cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of a mRNA molecule; and wherein the functional nucleic acid molecule results in overexpression of the EGF motif repeat protein protein; and (B) growing the transformed cell.

The present invention further provides a method of producing a cell containing reduced levels of an EGF motif repeat protein comprising: (A) transforming the cell with a functional nucleic acid molecule, wherein the functional nucleic acid molecule comprises a promoter region, wherein the promoter region is linked to a structural region, wherein the structural region comprises a nucleic acid molecule having a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or complements thereof; wherein the structural region is linked to a 3' non-translated sequence that functions in the cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of a mRNA molecule; and wherein the functional nucleic acid molecule results in co-suppression of the EGF motif repeat protein; and (B) growing the transformed cell.

The present invention further provides a method for reducing expression of an EGF motif repeat protein in a cell comprising: (A) transforming the cell with a nucleic acid molecule, the nucleic acid molecule having an exogenous promoter region which functions in a cell to cause the production of a mRNA molecule, wherein the exogenous promoter region is linked to a transcribed nucleic acid molecule having a transcribed strand and a non-transcribed strand, wherein the transcribed strand is complementary to a nucleic acid molecule having a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or complements thereof and the transcribed strand is complementary to an endogenous mRNA molecule; and wherein the transcribed nucleic acid molecule is linked to a 3' non-translated sequence that functions in the cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of a mRNA molecule; and (B) growing the transformed cell.

The present invention also provides a method of isolating a nucleic acid that encodes an EGF motif repeat protein or fragment thereof comprising: (A) incubating under conditions permitting nucleic acid hybridization, a first nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or complements thereof with a complementary second nucleic acid molecule obtained from a cell; (B) permitting

hybridization between the first nucleic acid molecule and the second nucleic acid molecule obtained from the cell; and (C) isolating the second nucleic acid molecule.

The present invention also provides an EGF motif repeat protein of SEQ ID NO:7, SEQ ID NO:8, variants and fragments thereof.

The present invention also provides a method for proliferating or expanding a hematopoietic stem cell population *ex vivo* comprising culturing the hematopoietic stem cell population in the presence of a cocktail of cytokines wherein the cocktail comprises a EGF motif repeat protein of the present invention, an IL-3 variant taught in WO 94/12639 and WO 94/12638, a fusion protein taught in WO 95/21197 and WO 95/21254, a G-CSF receptor agonists disclosed in WO 97/12977, a c-mpl receptor agonists disclosed in WO 97/12978, a IL-3 receptor agonists disclosed in WO 97/12979 and a multi-functional receptor agonists taught in WO 97/12985, flt3 receptor agonists disclosed in WO 98/18923, stem cell factor receptor agonists disclosed in WO 98/18924, erythropoietin receptor agonists disclosed in WO 98/18926, and multi-functional chimeric hematopoietic receptor agonists disclosed in WO 98/17810, and wherein the culturing promotes the growth of myeloid-type progenitor cells. As used herein "IL-3 variants" refer to IL-3 variants taught in WO 94/12639 and WO 94/12638. As used herein "fusion proteins" refer to fusion protein taught in WO 95/21197, and WO 95/21254. As used herein "G-CSF receptor agonists" refer to G-CSF receptor agonists disclosed in WO 97/12978. As used herein "c-mpl receptor agonists" refer to c-mpl receptor agonists disclosed in WO 97/12978. As used herein "IL-3 receptor agonists" refer to IL-3 receptor agonists disclosed in WO 97/12979. As used herein "multi-functional receptor agonists" refer to multi-functional receptor agonists taught in WO 97/12985. As used herein "flt3 receptor agonists" refers to flt3 receptor agonists disclosed in WO 98/18923. As used herein "stem cell factor receptor agonists" refers to stem cell factor receptor agonists disclosed in WO 98/18924. As used herein "erythropoietin receptor agonists" refers to erythropoietin receptor agonists disclosed in WO 98/18926. As used herein "multi-functional chimeric hematopoietic receptor

agonists" refers to multi-functional chimeric hematopoietic receptor agonists disclosed in WO 98/17810.

Definitions

The following is a list of abbreviations and the corresponding meanings as used interchangeably herein:

DMEM = Dulbecco's modified Eagle media

DTT = dithiothreitol

HBSS = Hanks balanced salt solution

HPLC = high performance liquid chromatography

IMDM = Iscove's modified Dulbecco's media

mg = milligram

ml = milliliter

mL = milliliter

μg = microgram

μl = microliter

PBS = phosphate buffered saline

TFA = trifluoracetic acid

ug = microgram

ul = microliter

The following is a list definitions of various terms used herein:

The term "altered" means that expression differs from the expression response of cells or tissues not exhibiting the phenotype.

The term "amino acid(s)" means all naturally occurring L-amino acids.

The term "chromosome walking" means a process of extending a genetic map by successive hybridization steps.

The term "cluster" means that BLAST scores from pairwise sequence comparisons of the member clones are similar enough to be considered identical with experimental error.

The term "complete complementarity" means that every nucleotide of one molecule is complementary to a nucleotide of another molecule.

The term "degenerate" means that two nucleic acid molecules encode for the same amino acid sequences but comprise different nucleotide sequences.

The term "exogenous genetic material" means any genetic material, whether naturally occurring or otherwise, from any source that is capable of being inserted into any organism.

The term "expansion" means the differentiation and proliferation of cells.

The term "expressed sequence tags (ESTs) means randomly sequenced members of a cDNA or complementary DNA library.

The term "expression response" means the mutation affecting the level or pattern of the expression encoded in part or whole by one or more nucleic acid molecules.

The term "fragment" means a nucleic acid molecule whose sequence is shorter than the target or identified nucleic acid molecule and having the identical, the substantial complement, or the substantial homologue of at least 10 contiguous nucleotides of the target or identified nucleic acid molecule.

The term "fusion molecule" means a protein-encoding molecule or fragment that upon expression, produces a fusion protein.

The term "fusion protein" means a protein or fragment thereof that comprises one or more additional peptide regions not derived from that protein.

The term "marker nucleic acid" means a nucleic acid molecule that is utilized to determine an attribute or feature (*e.g.*, presence or absence, location, correlation, etc.) of a molecule, cell, or tissue.

The term "mimetic compound" means a chemically synthesized compound with similar properties to a naturally occurring compound or a fragment of that compound, which exhibits an ability to specifically bind to antibodies directed against that compound.

The term "phenotype" means any of one or more characteristics of an organism, tissue, or cell.

The term "probe" means an agent that is utilized to determine an attribute or feature (e.g. presence or absence, location, correlation, etc.) of a molecule, cell, tissue, or organism.

The term "protein fragment" means a peptide or polypeptide molecule whose amino acid sequence comprises a subset of the amino acid sequence of that protein.

The term "protein molecule/peptide molecule" means any molecule that comprises five or more amino acids.

The term "recombinant" means any agent (e.g., DNA, peptide, etc.), that is, or results from, however indirectly, human manipulation of a nucleic acid molecule.

The term "selectable or screenable marker genes" means genes who's expression can be detected by a probe as a means of identifying or selecting for transformed cells.

The term "singleton" means a single clone.

The term "specifically bind" means that the binding of an antibody or peptide is not competitively inhibited by the presence of non-related molecules.

The term "specifically hybridizing" means that two nucleic acid molecules are capable of forming an anti-parallel, double-stranded nucleic acid structure.

The term "stem cell" means totipotent hematopoietic stem cells, as well as early precursors and progenitor cells, which can be isolated from bone marrow, spleen, or peripheral blood.

The term "stem cell growth factor" means a factor that acts on progenitor cells as well as stem cells.

The term "substantial complement" means that a nucleic acid sequence shares at least 80% sequence identity with the complement.

The term "substantial fragment" means a fragment that comprises at least 100 nucleotides.

The term "substantial homologue" means that a nucleic acid molecule shares at least 80% sequence identity with another.

The term "substantially hybridizing" means that two nucleic acid molecules can form an anti-parallel, double-stranded nucleic acid structure under conditions (e.g. salt and temperature) that permit hybridization of sequences that exhibit 90%

sequence identity or greater with each other and exhibit this identity for at least a contiguous 50 nucleotides of the nucleic acid molecules.

The term "substantially purified" means that one or more molecules that are or may be present in a naturally occurring preparation containing the target molecule will have been removed or reduced in concentration.

The term "tissue sample" means any sample that comprises more than one cell.

The term "variant" means a protein having a deletion and/or insertion, and/or substitution of at least one amino acid compared to the native amino acid sequence.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows Northern blot analysis of HS-5 poly A+ RNA to determine the transcript size of various candidate clones identified as potential growth factors. Poly A+ RNA was electrophoresed on formaldehyde gel and probed with Digoxigenin-labeled PCR fragment of the EST of interest. The RNA marker sizes are shown in numbers.

Figure 2 shows the DNA sequence SEQ ID NO:3 that encodes the deduced open reading frame of the EGF motif repeat protein of SEQ ID NO:7. The arrow indicates potential cleavage sites between residues 34 and 35.

Figure 3 shows the secretion signal, the 10 EGF motif repeats and the potential transmembrane domain of the EGF motif repeat protein of SEQ ID NO:7.

Figure 4 shows the alignment of the 10 EGF motif repeats from SEQ ID NO:7 with the EGF motif consensus using HMMER 2.1.1 (Dec 1998) software Copyright © 1992-1998 Washington University School of Medicine HMMER is freely distributed under the GNU General Public License (GPL).

DETAILED DESCRIPTION OF THE INVENTION

A library of expressed sequence tags (ESTs) was created from a cDNA library derived from the HS-5 cell line. ESTs are randomly sequenced members of a cDNA library (or complementary DNA)(McCombie *et al., Nature Genetics 1:*124-130 (1992); Kurata *et al., Nature Genetics 8:* 365-372 (1994); Okubo *et al., Nature Genetics 2:* 173-179 (1992), all of which references are incorporated in their entirety). The selected clones comprise inserts that can represent a copy of up to the full length of an mRNA transcript.

The cDNA library was prepared in part by subtracting the mRNA in HS-5 that was common to HS-21, HS-27, and HS-23. The EST sequences were compared to sequences in various databases by computer algorithms. The sequences set forth herein importantly have a growth factor and/or secretion motif. Accordingly, these sequences correspond to the EGF motif repeat protein expressed by HS-5, and found in HS-5 conditioned media, but not expressed by HS-21, HS-27, and HS-23.

The degeneracy of the genetic code, which allows different nucleic acid sequences to code for the same protein or peptide, is known in the literature (U.S. Patent No. 4,757,006). As used herein, a nucleic acid molecule is "degenerate" of another nucleic acid molecule when the nucleic acid molecules encode for the same amino acid sequences but comprise different nucleotide sequences. Nucleic acid molecules of the present invention include, but are not limited to, homologous nucleic acid molecules that are degenerate of those set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or complements thereof.

As used herein, a "substantial homologue" of a nucleic acid molecule is one that shares at least 80% sequence identity therewith. Variations are due to the degeneracy of the genetic code. Preferably, a homologue shares at least 90%, more preferably at least 95%, sequence identity with the target nucleic acid molecule. In some embodiments, the substantial homologue will differ from the target nucleic acid molecule by no more than 5 nucleotides, preferably no more than 3 nucleotides.

A nucleic acid molecule is said to be the "complement" of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to a nucleotide of the other (complete complementarity). A "substantial complement" shares at least 80% sequence identity with the complement. Preferably, the

substantial complement shares at least 90%, more preferably at least 95%, sequence identity with the complement. In some embodiments, the substantial complement will differ from the complement by no more than 5 nucleotides, preferably no more than 3 nucleotides.

A "fragment" as used herein means a nucleic acid molecule whose sequence is shorter than the target or identified nucleic acid molecule and having the identical, the substantial complement, or the substantial homologue of at least 10 contiguous nucleotides of the target or identified nucleic acid molecule. Accordingly, a fragment contains at least 10 nucleotides, typically at least 50 nucleotides, more typically at least 60 nucleotides, and preferably at least 100 nucleotides. The upper limit on the number of nucleotides is essentially only limited by the number of nucleotides in the target nucleic acid molecule.

As indicated above, the fragment can be of a substantial homologue or substantial complement of one of the sequences set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and SEQ ID NO:4. Preferably, the fragment is identical or complementary to at least 50 contiguous nucleotides in one of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and SEQ ID NO:4. Such a "substantial fragment" preferably comprises at least 100 nucleotides.

As used herein, the term "substantially purified," means that one or more molecules that are or may be present in a naturally occurring preparation containing the target molecule will have been removed or reduced in concentration.

It is understood that the sequences of the present invention described above, including the homologues, complements, and fragments, also include the labeled forms thereof (e.g., fluorescent labels (Prober *et al., Science* 238:336-340 (1987); Albarella *et al.,* European Patent 144914), chemical labels (Sheldon *et al.,* U.S. Patent 4,582,789; Albarella *et al.,* U.S. Patent 4,563,417), modified nucleotides (Miyoshi *et al.,* European Patent 119448)).

The nucleic acid molecules comprising SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or complements thereof, substantial homologues thereof, and substantial fragments thereof, can encode, either by themselves or as part of a longer sequence, an EGF motif repeat protein, or fragment thereof. Accordingly, one aspect of the present invention is a substantially purified nucleic acid molecule

that comprises at least one nucleic acid sequence that is identical to, a substantial homologue to, or a substantial complement to a sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4. In one embodiment of the present invention, one or more of the nucleic acid molecules of the present invention share between 90% and 100% sequence identity with one or more of the nucleic acid sequences set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or complements thereof. In a further aspect of the present invention, one or more of the nucleic acid molecules share between 95% and 100%, preferably between 98% and 100%, and more preferably between 99% and 100% sequence identity with one or more of the nucleic acid sequences set forth in SEQ ID NO: 1 through SEQ ID NO:297 or complements thereof.

In a preferred embodiment, a nucleic acid of the present invention will specifically hybridize to one or more of the nucleic acid molecules set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or complements thereof under moderately stringent conditions, for example at about 2x sodium chloride/sodium citrate (SSC) and about 40°C. In a particularly preferred embodiment, a nucleic acid of the present invention will specifically hybridize to one or more of the nucleic acid molecules set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or complements thereof under high stringency conditions.

Conventional stringency conditions are described by Sambrook *et al.*, In: *Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press,* Cold Spring Harbor, New York (1989)), and by Haymes *et al.*, In: *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, DC (1985). Departures from complete complementarity are therefore permissible, as long as such departures do not completely preclude the capacity of the molecules to form a double-stranded structure.

Appropriate stringency conditions which promote DNA hybridization, for example, 6x SCC at about 45°C, followed by a wash of 2x SSC at 50°C, are known to those skilled in the art or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6). For example, the salt concentration in the wash step can be selected from a low stringency of about 2x

SSC at 50°C to a high stringency of about 0.2x SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C. Both temperature and salt may be varied, or either the temperature or the salt concentration may be held constant while the other variable is changed.

Nucleic acid molecules or fragments thereof are capable of specifically hybridizing to other nucleic acid molecules under certain circumstances. As used herein, two nucleic acid molecules are said to be capable of "specifically hybridizing" to one another if the two molecules are capable of forming an anti-parallel, double-stranded nucleic acid structure.

The present invention also relates to the substantially purified protein, fragment thereof, or polypeptide molecule encoded by the above-mentioned sequences. That is, substantially purified recombinant proteins, protein fragments and polypeptides. As used herein, the term "recombinant" means any agent (e.g., DNA, peptide, etc.), that is, or results from, however indirectly, human manipulation of a nucleic acid molecule. As used herein, the term "protein molecule" or "peptide molecule" includes any molecule that comprises five or more amino acids. It is well known in the art that proteins may undergo modification, including post-translational modifications, such as, but not limited to, disulfide bond formation, glycosylation, phosphorylation, or oligomerization. Thus, as used herein, the term "protein molecule" or "peptide molecule" includes any protein molecule that is modified by any biological or non-biological process. The terms "amino acid" and "amino acids" refer to all naturally occurring L-amino acids. This definition is meant to include norleucine, ornithine, homocysteine, and homoserine.

Non-limiting examples of the protein or fragment molecules of the present invention are a novel EGF motif repeat protein or fragment thereof encoded by a sequence that substantially hybridizes with a sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or complements thereof. Preferably, the protein is an EGF motif repeat protein that is encoded by the DNA sequence of SEQ ID NO:2, SEQ ID NO:3 or SEQ ID NO:4 or a fragment thereof. Another aspect of the nucleic acid molecules of the present invention is that they can encode a homologue or fragment thereof of a EGF motif repeat protein. As used herein, a

homologue protein molecule or fragment thereof is a counterpart protein molecule or fragment thereof in a second species (*e.g.*, human's interleukin-2 cytokine is a homologue of mouse's interleukin-2 cytokine).

An embodiment of the present invention is the EGF motif repeat protein shown in Figure 3. The EGF motif repeat protein sequence, SEQ ID NO:7, is deduced from DNA sequence of SEQ ID NO:2. The EGF protein has a secretion leader sequence, 10 EGF repeats, a potential transmembrane domain and an intracellular region (Figure 3). The EGF repeats 1 and 2 are separated from EGF repeats 3-10 by 172 amino acids. Another embodiment of the present invention is the mature EGF motif repeat protein of SEQ ID NO:8, which has secretion leader sequence cleaved at potential cleavage sites between residues 34 and 35. Another embodiment of the present is a variant of the EGF motif repeat protein of SEQ ID NO:7 or SEQ ID NO:8 in which at least one amino acid is deleted, and/or inserted, and/or substituted compared to the native amino acid sequence.

One or more of the protein or fragment of peptide molecules may be produced by chemical synthesis, or more preferably, by expressing in a suitable bacterial or eukaryotic host. Suitable methods for expression are described by Sambrook *et al.*, (In: *Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press*, Cold Spring Harbor, New York (1989)), or similar texts.

A "protein fragment" is a peptide or polypeptide molecule whose amino acid sequence comprises a subset of the amino acid sequence of that protein. A protein or fragment thereof that comprises one or more additional peptide regions not derived from that protein is a "fusion protein". Such molecules may be derivatized to contain carbohydrate or other moieties (such as keyhole limpet hemocyanin, etc.). Fusion proteins or peptide molecules of the present invention are preferably produced by recombinant means.

The sequences of the present invention can be formed by well-known and conventional techniques. For example, genetic engineering techniques may be employed in the construction of the DNA sequences of the present invention (U.S Patent No. 4,935,233; Sambrook *et al.*, "*Molecular Cloning A Laboratory Manual*", Cold Spring Harbor (1989)). Any of a variety of methods may be used to obtain one or more of the above-described nucleic acid molecules (Zamechik *et al.*, *Proc. Natl.*

Acad. Sci. (U.S.A.) 83:4143-4146 (1986), the entirety of which is incorporated by reference; Goodchild *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 85*:5507-5511 (1988), the entirety of which is incorporated by reference; Wickstrom *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 85*:1028-1032 (1988), the entirety of which is incorporated by reference; Holt *et al.*, *Molec. Cell. Biol. 8*:963-973 (1988), the entirety of which is incorporated by reference; Gerwirtz *et al.*, *Science 242*:1303-1306 (1988), the entirety of which is incorporated by reference; Anfossi *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 86*:3379-3383 (1989), the entirety of which is incorporated by reference; Becker *et al.*, *EMBO J. 8*:3685-3691 (1989); the entirety of which is incorporated by reference). Automated nucleic acid synthesizers may be employed for this purpose. In lieu of such synthesis, the disclosed nucleic acid molecules may be used to define a pair of primers that can be used with the polymerase chain reaction (Mullis *et al.*, *Cold Spring Harbor Symp. Quant. Biol. 51*:263-273 (1986); Erlich *et al.*, European Patent 50,424; European Patent 84,796, European Patent 258,017, European Patent 237,362; Mullis, European Patent 201,184; Mullis *et al.*, U.S. Patent 4,683,202; Erlich, U.S. Patent 4,582,788; and Saiki, R. *et al.*, U.S. Patent 4,683,194, all of which are incorporated by reference in their entirety) to amplify and obtain any desired nucleic acid molecule or fragment.

The nucleic acid molecules of the present invention may be used as probes in connection with methods that require probes. As used herein, a "probe" is a nucleic acid molecule that is utilized to determine an attribute or feature (*e.g.*, presence or absence, location, correlation, etc.) of a molecule, cell, or tissue. Preferably, the nucleic acids of the present invention are used as a probe for a stem cell.

Nucleic acid molecules and fragments thereof of the present invention may be employed to obtain other nucleic acid molecules from the same species (e.g., ESTs from the HS-5 cell line may be utilized to obtain other nucleic acid molecules from other hematopoietic stem or stromal cell lines). Such nucleic acid molecules include the nucleic acid molecules that encode, the complete coding sequence of a protein, and promoters and flanking sequences of such molecules. In addition, such nucleic acid molecules include nucleic acid molecules that encode for other isozymes or gene family members. Such molecules can be readily obtained by using the above-described nucleic acid molecules or fragments thereof to screen cDNA or

genomic libraries obtained from a hematopoietic stem cell. Methods for forming such libraries are well known in the art.

Nucleic acid molecules and fragments thereof of the present invention may also be employed to obtain nucleic acid homologues. Such homologues include the nucleic acid molecule of other hematopoietic cell lines or other tissues (*e.g.*, HS-5 stromal cell line, etc.) including the nucleic acid molecules that encode, in whole or in part, protein homologues of other hematopoietic stem cell lines or other tissues, sequences of genetic elements such as promoters and transcriptional regulatory elements. Such molecules can be readily obtained by using the above-described nucleic acid molecules or fragments thereof to screen cDNA or genomic libraries. Methods for forming such libraries are well known in the art. Such homologue molecules may differ in their nucleotide sequences from those found in one or more of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or complements thereof because complete complementarity is not needed for stable hybridization. The nucleic acid molecules of the present invention therefore also include molecules that, although capable of specifically hybridizing with the nucleic acid molecules, may lack "complete complementarity."

Promoter sequence(s) and other genetic elements, including but not limited to transcriptional regulatory flanking sequences, associated with one or more of the disclosed nucleic acid sequences can also be obtained using the disclosed nucleic acid sequence provided. In one embodiment, such sequences are obtained by incubating EST nucleic acid molecules or preferably fragments thereof with members of genomic libraries (*e.g.* HS-5 stromal cell line) and recovering clones that hybridize to the EST nucleic acid molecule or fragment thereof. In a second embodiment, methods of "chromosome walking," or inverse PCR may be used to obtain such sequences (Frohman *et al., Proc. Natl. Acad. Sci. (U.S.A.) 85*:8998-9002 (1988); Ohara *et al., Proc. Natl. Acad. Sci. (U.S.A.) 86*: 5673-5677 (1989); Pang *et al., Biotechniques 22(6):* 1046-1048 (1977); Huang *et al., Methods Mol. Biol. 69:* 89-96 (1997); Huang, *et al., Method Mol. Biol. 67:*287-294 (1997); Benkel *et al., Genet. Anal. 13:* 123-127 (1996); Hartl *et al., Methods Mol. Biol. 58:* 293-301 (1996), all of which are incorporated by reference in their entirety).

The nucleic acid molecules of the present invention may be used to isolate promoters of cell-enhanced, cell-specific, tissue-enhanced, tissue-specific, and developmentally- or physiologically regulated expression profiles. Isolation and functional analysis of the 5' flanking promoter sequences of these genes from genomic libraries, for example, using genomic screening methods and PCR techniques would result in the isolation of useful promoters and transcriptional regulatory elements. These methods are known to those of skill in the art and have been described (see, for example, Birren *et al.*, *Genome Analysis: Analyzing DNA*, 1, (1997), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., the entirety of which is incorporated by reference). Promoters obtained utilizing the nucleic acid molecules of the present invention could also be modified to affect their control characteristics. Examples of such modifications would include, but are not limited to, enhancer sequences as reported by Kay *et al.*, *Science 236*:1299 (1987). Genetic elements such as these could be used to enhance gene expression of new and existing hematopoietic stem cell growth factors.

In one aspect of the present invention, one or more of the nucleic molecules are used to determine whether a cell (preferably a hematopoietic cell) has a mutation affecting the level (i.e., the concentration of mRNA in a sample, etc.) or pattern (i.e., the kinetics of expression, rate of decomposition, stability profile, etc.) of the expression encoded in part or whole by one or more of the nucleic acid molecules of the present invention (collectively, the "Expression Response" of a cell or tissue). As used herein, the Expression Response manifested by a cell or tissue is said to be "altered" if it differs from the Expression Response of cells or tissues not exhibiting the phenotype. To determine whether an Expression Response is altered, the Expression Response manifested by the cell or tissue exhibiting the phenotype is compared with that of a similar cell or tissue sample not exhibiting the phenotype. It is not necessary to re-determine the Expression Response of the cell or tissue sample not exhibiting the phenotype each time such a comparison is made; the Expression Response of a particular cell, for example, may be compared with previously obtained values of normal cells. As used herein, the "phenotype" of the organism is any of one or more characteristics of an organism, tissue, or cell (e.g., cell growth, cell differentiation, etc.). A change in genotype or phenotype may be

transient or permanent. Also as used herein, a "tissue sample" is any sample that comprises more than one cell. In one aspect, a tissue sample comprises cells that share a common characteristic (e.g., derived from hematopoietic stem cell line, etc.).

In one embodiment of the present invention, an evaluation can be conducted to determine whether a particular mRNA molecule is present. One or more of the nucleic acid molecules of the present invention, preferably one or more of the EST nucleic acid molecules of the present invention are utilized to detect the presence or quantity of the mRNA species. Such molecules are then incubated with cell or tissue extracts of a cell under conditions sufficient to permit nucleic acid hybridization. The detection of double-stranded probe-mRNA hybrid molecules is indicative of the presence of the mRNA. The amount of hybrid molecules formed is proportional to the amount of mRNA. Thus, such probes may be used to ascertain the level and extent of the mRNA production in a cell or tissue. Such nucleic acid hybridization may be conducted under quantitative conditions (thereby providing a numerical value of the amount of the mRNA present). Alternatively, the assay may be conducted as a qualitative assay indicating the presence of the mRNA, or that its level exceeds a predefined value.

A principle of *in situ* hybridization is that a labeled, single-stranded nucleic acid probe will hybridize to a complementary strand of cellular DNA or RNA and, under the appropriate conditions, these molecules will form a stable hybrid. When nucleic acid hybridization is combined with histological techniques, specific DNA or RNA sequences can be identified within a single cell. An advantage of *in situ* hybridization over more conventional techniques for the detection of nucleic acids is that it allows an investigator to determine the precise local concentration of the nucleic acid molecule within a tissue or cell (Angerer *et al., Dev. Biol. 101*: 477-484 (1984); Angerer *et al., Dev. Biol. 112*: 157-166 (1985); Dixon *et al., EMBO J. 10*: 1317-1324 (1991)). *In situ* hybridization may be used to measure the steady-state level of RNA accumulation. It is a sensitive technique and RNA sequences present in as few as 5-10 copies per cell can be detected (Hardin *et al., J. Mol. Biol. 202*: 417-431 (1989)). A number of protocols have been devised for *in situ* hybridization, each with tissue preparation, hybridization, and washing conditions.

In situ hybridization also allows for the localization of proteins within a tissue or cell (Wilkinson, *In Situ Hybridization,* Oxford University Press, Oxford (1992)). It is understood that one or more of the molecules of the present invention, preferably one or more of the EST nucleic acid molecules of the present invention or one or more of the antibodies of the present invention, may be utilized to detect the level or pattern of a EGF motif repeat protein pathway enzyme or mRNA thereof by *in situ* hybridization.

Fluorescent *in situ* hybridization allows the localization of a particular DNA sequence along a chromosome. This technique is useful for gene mapping, following chromosomes in hybrid lines or detecting chromosomes with translocations, transversions or deletions, among others. It is understood that the nucleic acid molecules of the present invention may be used as probes or markers to localize sequences along a chromosome.

A microarray-based method for high-throughput monitoring of gene expression may be utilized to measure gene-specific hybridization targets. This 'chip'-based approach involves using microarrays of nucleic acid molecules as gene-specific hybridization targets to quantitatively measure expression of the corresponding genes (Schena *et al.*, *Science 270*:467-470 (1995); Shalon, Ph.D. Thesis. Stanford University (1996)). Every nucleotide in a large sequence can be queried at the same time. Hybridization can be used to efficiently analyze nucleotide sequences.

Several microarray methods have been described. One method compares the sequences to be analyzed by hybridization to a set of oligonucleotides representing all possible subsequences (Bains and Smith, *J. Theor. Biol. 135*:303-307 (1989)). A second method hybridizes the sample to an array of oligonucleotide or cDNA molecules. An array consisting of oligonucleotides complementary to subsequences of a target sequence can be used to determine the identity of a target sequence, measure its amount, and detect differences between the target and a reference sequence. Nucleic acid molecule microarrays may also be screened with protein molecules, or fragments thereof, to determine nucleic acid molecules that specifically bind protein molecules, or fragments thereof.

The microarray approach may be used with polypeptide targets (U.S. Patent No. 5,445,934; U.S. Patent No: 5,143,854; U.S. Patent No. 5,079,600; U.S. Patent No. 4,923,901). Essentially, polypeptides are synthesized on a substrate (microarray) and these polypeptides can be screened with either protein molecules or fragments thereof including but not limited to antibodies, or nucleic acid molecules in order to screen for either protein molecules or fragments thereof, or nucleic acid molecules that specifically bind the target polypeptides, or small molecules with substantial affinity for protein molecules, or small molecules with substantial affinity for specific protein molecules. Implementation of these techniques relies on recently developed combinatorial technologies to generate any ordered array of a large number of oligonucleotide probes (Fodor *et al.*, *Science 251*:767-773 (1991)). It is understood that one or more of the nucleic acid molecules or protein or fragments thereof of the present invention may be utilized in a microarray-based method.

Site-directed mutagenesis may be utilized to modify nucleic acid sequences. This technique allows one or more of the amino acids encoded by a nucleic acid molecule to be altered (e.g. a threonine to be replaced by a methionine). Three basic methods for site-directed mutagenesis are often employed. These are cassette mutagenesis (Wells *et al.*, *Gene 34*:315-323 (1985)), primer extension (Gilliam *et al.*, *Gene 12*:129-137 (1980); Zoller and Smith, *Methods Enzymol. 100*:468-500 (1983); Dalbadie-McFarland *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 79*:6409-6413 (1982)) and methods based upon PCR (Scharf *et al.*, *Science 233*:1076-1078 (1986); Higuchi *et al.*, *Nucleic Acids Res. 16*:7351-7367 (1988)). Site-directed mutagenesis approaches are also described in European Patent 0 385 962, European Patent 0 359 472, and PCT Patent Application WO 93/07278.

Site-directed mutagenesis strategies applied for both *in vitro* as well as *in vivo* site directed mutagenesis have been reviewed by Ling and Robinson, *Anal. Biochem. 254*:157-178 (1997); Chong *et al.*, *Biotechniques 17*:719-720 (1994); Weiner *et al.*, *Gene 151*:119-123 (1994); Chen *et al.*, *Nucleic Acids Res. 25*:682-684 (1997); Weiner *et al.*, *Gene 126*:35-41 (1993).

Any of the nucleic acid molecules of the present invention may either be modified by site-directed mutagenesis or used as, for example, nucleic acid

molecules that are used to target other nucleic acid molecules for modification. Mutants containing one or more altered nucleotide can be constructed using common laboratory techniques such as isolating restriction fragments and ligating such fragments into an expression vector (*see,* for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989)).

One aspect of the present invention concerns antibodies, single-chain antigen binding molecules, or other proteins that specifically bind to one or more of the protein or peptide molecules of the present invention, and their homologues, fusions or fragments. Such antibodies may be used to quantitatively or qualitatively detect the protein or peptide molecules of the present invention. As used herein, an antibody or peptide is said to "specifically bind" to a protein or peptide molecule of the present invention if such binding is not competitively inhibited by the presence of non-related molecules.

Nucleic acid molecules that encode all or part of the protein of the present invention can be expressed, by recombinant means, to yield protein or peptides that can be used to elicit antibodies that are capable of binding the expressed protein or peptide. Such antibodies may be used in immunoassays for that protein. Such protein-encoding molecules, or their fragments may be a "fusion molecule" (i.e., a part of a larger nucleic acid molecule) such that, upon expression, produce a fusion protein. It is understood that any of the nucleic acid molecules of the present invention may be expressed, by recombinant means, to yield proteins or peptides encoded by these nucleic acid molecules.

The antibodies that specifically bind proteins and protein fragments of the present invention may be polyclonal or monoclonal, and may comprise intact immunoglobulins, or antigen binding portions of immunoglobulins fragments (such as (F(ab'), $F(ab')_2$), or single-chain immunoglobulins producible, for example, by recombinant means. It is understood that practitioners are familiar with the standard resource materials, which describe specific conditions and procedures for the construction, manipulation, and isolation of antibodies (*see*, for example, Harlow and Lane, *In Antibodies: A Laboratory Manual,* Cold Spring Harbor Press, Cold Spring Harbor, New York (1988)).

Murine monoclonal antibodies are particularly preferred. BALB/c mice are preferred for this purpose, however, equivalent strains may also be used. The animals are preferably immunized with approximately 25 μg of purified protein (or fragment thereof) that has been emulsified in a suitable adjuvant (such as TiterMax adjuvant (Vaxcel, Norcross, GA)). Immunization is preferably conducted at two intramuscular sites, one intraperitoneal site, and one subcutaneous site at the base of the tail. An additional i.v. injection of approximately 25 μg of antigen is preferably given in normal saline three weeks later. After approximately 11 days following the second injection, the mice may be bled and the blood screened for the presence of anti-protein or peptide antibodies. Preferably, a direct binding Enzyme-Linked Immunoassay (ELISA) is employed for this purpose.

More preferably, the mouse having the highest antibody titer is given a third i.v. injection of approximately 25 μg of the same protein or fragment. The splenic leukocytes from this animal may be recovered three days later, and then permitted to fuse, most preferably, using polyethylene glycol, with cells of a suitable myeloma cell line (such as, for example, the P3X63Ag8.653 myeloma cell line). Hybridoma cells are selected by culturing the cells under "HAT" (hypoxanthine-aminopterin-thymine) selection for about one week. The resulting clones may then be screened for their capacity to produce monoclonal antibodies (mAbs), preferably by direct ELISA.

In one embodiment, anti-protein or peptide monoclonal antibodies are isolated using a fusion of a protein or peptide of the present invention, or conjugate of a protein or peptide of the present invention, as immunogens. Thus, for example, a group of mice can be immunized using a fusion protein emulsified in Freund's complete adjuvant (*e.g.* approximately 50 μg of antigen per immunization). At three-week intervals, an identical amount of antigen is emulsified in Freund's incomplete adjuvant and used to immunize the animals. Ten days following the third immunization, serum samples are taken and evaluated for the presence of antibody. If antibody titers are too low, a fourth booster can be employed. Polysera capable of binding the protein or peptide can also be obtained using this method.

In a preferred procedure for obtaining monoclonal antibodies, the spleens of the above-described immunized mice are removed, disrupted, and immune

splenocytes are isolated over a ficoll gradient. The isolated splenocytes are fused, using polyethylene glycol with BALB/c-derived HGPRT (hypoxanthine guanine phosphoribosyl transferase) deficient P3x63xAg8.653 plasmacytoma cells. The fused cells are plated into 96 well microtiter plates and screened for hybridoma fusion cells by their capacity to grow in culture medium supplemented with hypothanthine, aminopterin, and thymidine for approximately 2-3 weeks.

Hybridoma cells that arise from such incubation are preferably screened for their capacity to produce an immunoglobulin that binds to a protein of interest. An indirect ELISA may be used for this purpose. In brief, the supernatants of hybridomas are incubated in microtiter wells that contain immobilized protein. After washing, the titer of bound immunoglobulin can be determined using, for example, a goat anti-mouse antibody conjugated to horseradish peroxidase. After additional washing, the amount of immobilized enzyme is determined (for example through the use of a chromogenic substrate). Such screening is performed as quickly as possible after the identification of the hybridoma in order to ensure that a desired clone is not overgrown by non-secreting neighbor cells. Desirably, the fusion plates are screened several times since the rates of hybridoma growth vary. In a preferred sub-embodiment, a different antigenic form may be used to screen the hybridoma. Thus, for example, the splenocytes may be immunized with one immunogen, but the resulting hybridomas can be screened using a different immunogen. It is understood that any of the protein or peptide molecules of the present invention may be used to raise antibodies.

As discussed below, such antibody molecules or their fragments may be used for diagnostic purposes. Where the antibodies are intended for diagnostic purposes, it may be desirable to derivatize them, for example with a ligand group (such as biotin) or a detectable marker group (such as a fluorescent group, a radioisotope or an enzyme).

The ability to produce antibodies that bind the protein or peptide molecules of the present invention permits the identification of mimetic compounds of those molecules. A "mimetic compound" is a compound that is closely related to a compound in either structure or function, or a fragment of that compound, but which

nonetheless exhibits an ability to specifically bind to antibodies directed against that compound.

It is understood that any of the antibodies of the present invention may be expressed and that such expression can result in a physiological effect. It is also understood that any of the expressed antibodies may be catalytic.

Another aspect of the present invention provides plasmid DNA vectors for use in the expression of the EGF motif repeat protein of the present invention. These vectors contain the DNA sequences described above which code for the polypeptides of the invention. Appropriate vectors which can transform microorganisms capable of expressing the EGF motif repeat protein include expression vectors comprising nucleotide sequences coding for the EGF motif repeat protein joined to transcriptional and translational regulatory sequences which are selected according to the host cells used.

Vectors incorporating modified sequences as described above are included in the present invention and are useful in the production of the EGF motif repeat proteins. The vector employed in the method also contains selected regulatory sequences in operative association with the DNA coding sequences of the invention and which are capable of directing the replication and expression thereof in selected host cells.

Transfer of a nucleic acid that encodes for a protein can result in overexpression of that protein in a transformed cell. One or more of the proteins or fragments thereof encoded by nucleic acid molecules of the present invention may be overexpressed in a transformed cell. Particularly, any of the hematopoietic stem cell growth factors or fragments thereof may be overexpressed in a transformed cell. Such overexpression may be the result of transient or stable transfer of the exogenous genetic material. "Exogenous genetic material" is any genetic material, whether naturally occurring or otherwise, from any source that is capable of being inserted into any organism.

A construct or vector may include a promoter to express the protein or protein fragment of choice. The promoter of the present invention may be a hematopoietic stem cell-specific promoter. Preferably, the hematopoietic stem cell-specific promoter of the present invention is the CD34 promoter (Burn *et al.*, U.S.

Patent No. 5,556,954). Additional promoters that can be used in the present invention include the glucose-6-phosphatase promoter (Yoshiuchi *et al.*, *J. Clin. Endocrin. Metab. 83*:1016-1019 (1998)), interleukin-1 alpha promoter (Mori and Prager, *Leuk. Lymphoma 26*:421-433 (1997)), CMV promoter (Tong *et al.*, *Anticancer Res. 18*:719-725 (1998); Norman *et al.*, *Vaccine 15*:801-803 (1997)); RSV promoter (Elshami *et al.*, *Cancer Gene Ther. 4*:213-221 (1997); Baldwin *et al.*, *Gene Ther. 4*:1142-1149 (1997)); SV40 promoter (Harms and Splitter, *Hum. Gene Ther. 6*:1291-1297 (1995)), CD11c integrin gene promoter (Corbi and Lopez-Rodriguez, *Leuk. Lymphoma 25*:415-425 (1997)), GM-CSF promoter (Shannon *et al.*, *Crit. Rev. Immunol. 17*:301-323 (1997)), interleukin-5R alpha promoter (Sun *et al.*, *Curr. Top. Microbiol. Immunol 211*:173-187 (1996)), interleukin-2 promoter (Serfing *et al.*, *Biochim. Biophys. Acta 1263*:181-200 (1995); O'Neill *et al.*, *Transplant Proc. 23*:2862-2866 (1991)), c-fos promoter (Janknecht, *Immunobiology 193*:137-142 (1995); Janknecht *et al.*, *Carcinogenesis 16*:443-450 (1995); Takai *et al.*, *Princess Takamatsu Symp. 22*:197-204 (1991)), h-ras promoter (Rachal *et al.*, *EXS 64*:330-342 (1993)), and DMD gene promoter (Ray *et al.*, *Adv. Exp. Med. Biol. 280*:107-111 (1990)).

Promoters suitable for expression of the EGF motif repeat protein of the present invention in bacteria have been described by Hawley and McClure, *Nucleic Acids Res. 11*:2237-2255 (1983), and Harley and Reynolds, *Nucleic Acids Res. 15*:2343-2361 (1987). Such promoters include, for example, the rec A promoter (Fernandez de Henestrosa *et al.*, *FEMS Microbiol. Lett. 147*:209-213 (1997); Nussbaumer *et al.*, *FEMS Microbiol. Lett. 118*:57-63 (1994); Weisemann *et al.*, *Biochimie 73*: 457-470 (1991)), the ptac promoter (Hasan *et al.*, *Gene 56*:141-151 (1987); Marsh, *Nucleic Acids Res. 14*:3603 (1986)); and a ptac-rec A hybrid promoter.

It is preferred that the particular promoter selected is capable of causing sufficient expression to result in the production of an effective amount of the EGF motif repeat protein to cause the desired phenotype.

Constructs or vectors may also include with the coding region of interest a nucleic acid sequence that acts, in whole or in part, to terminate transcription of that region.

It is understood that the nucleic acid molecules of the present invention may be used to isolate regulatory elements preferentially associated with hematopoietic stem cell growth factors. For example, the nucleic acid molecules of the present invention may be used to isolate promoter sequences associated with hematopoietic stem cell growth factors. More preferably, the nucleic acid molecules of the present invention are used to isolate promoter sequences associated with the growth factor of the HS-5 cell line of the present invention.

Translational enhancers may also be incorporated as part of the vector DNA. DNA constructs could contain one or more 5' non-translated leader sequences that may serve to enhance expression of the gene products from the resulting mRNA transcripts. Such sequences may be derived from the promoter selected to express the gene or can be specifically modified to increase translation of the mRNA. Such regions may also be obtained from viral RNAs, from suitable eukaryotic genes, or from a synthetic gene sequence.

A vector or construct may also include a screenable marker. Screenable markers may be used to monitor expression. Exemplary screenable markers include β-glucuronidase encoded by the *uidA* gene (GUS) (Jefferson, *Plant Mol. Biol. Rep. 5:* 387-405 (1987); Jefferson *et al., EMBO J. 6:* 3901-3907 (1987)); β-lactamase (Sutcliffe *et al., Proc. Natl. Acad. Sci. (U.S.A.) 75*: 3737-3741 (1978)), luciferase (Clontech) (Ow *et al., Science 234*: 856-859 (1986)); β-galactosidase (Clontech); GST (Stratagene); Protein A (Calbiochem); blue fluorescent protein (Clontech); and green fluorescent protein (Clontech).

Included within the terms "selectable or screenable marker genes" are also genes that encode a secretable marker whose secretion can be detected as a means of identifying or selecting for transformed cells. Examples include markers, which encode a secretable antigen that can be identified by antibody interaction, or even secretable enzymes that can be detected by catalytic reactions. Secretable proteins fall into a number of classes, including small, diffusible proteins, which are detectable, (*e.g.,* by ELISA), small active enzymes, which are detectable in extracellular solution (*e.g.,* α-amylase, β-lactamase, phosphinothricin transferase), or proteins, which are inserted or trapped in the cell membrane (such as proteins

which include a leader sequence). Other possible selectable and/or screenable marker genes will be apparent to those of skill in the art.

As another aspect of the present invention, there is provided a method for producing the EGF motif repeat protein. Suitable cells or cell lines may be bacterial cells. For example, the various strains of *E. coli* are well known as host cells in the field of biotechnology. Examples of such strains include *E. coli* strains JM101 (Yanish-Perron *et al. Gene 33:*103-119 (1985)) and MON105 (Obukowicz *et al., Applied Environmental Microbiology 58:*1511-1523 (1992)). Also included in the present invention is the expression of the EGF motif repeat protein utilizing a chromosomal expression vector for *E. coli* based on the bacteriophage Mu (Weinberg *et al., Gene 126:*25-33 (1993)). Various strains of *B. subtilis* may also be employed in this method. Many strains of yeast cells known to those skilled in the art are also available as host cells for expression of the polypeptides of the present invention.

When expressed in the *E. coli* cytoplasm, the gene encoding the EGF motif repeat protein of the present invention may also be constructed such that the 5' end of the gene codons are added to encode Met^{-2}-Ala^{-1} - or Met^{-1} at the N-terminus of the protein. The N termini of proteins made in the cytoplasm of *E. coli* are affected by post-translational processing by methionine aminopeptidase (Bassat *et al., J. Bac. 169:*751-757 (1987)) and possibly by other peptidases so that upon expression the methionine is cleaved off the N-terminus. The EGF motif repeat protein of the present invention may be EGF motif repeat polypeptides having Met^{-1}, Ala^{-1} or Met^{-2}-Ala^{-1} at the N-terminus. These EGF motif repeat polypeptides may also be expressed in *E. coli* by fusing a secretion signal peptide of the N-terminus. This signal peptide can be cleaved from the polypeptide as part of the secretion process.

Under another embodiment, the EGF motif repeat protein of the present invention is expressed in a yeast cell, preferably *Saccharomyces cerevisiae*. The protein of the present invention can be expressed in *S. cerevisiae* by fusing it to the N-terminus of the URA3, CYC1 or ARG3 genes (Guarente and Ptashne, *Proc. Natl. Acad. Sci. (U.S.A.) 78:*2199-2203 (1981); Rose *et al., Proc. Natl. Acad. Sci. (U.S.A.) 78:*2460-2464 (1981); and Crabeel *et al., EMBO J. 2:*205-212 (1983)). Alternatively, the EGF motif repeat protein of the present invention can be fused to

either the PGK or TRP1 genes (Tuite *et al.*, *EMBO J. 1*:603-608 (1982); and Dobson *et al.*, *Nucleic Acids. Res. 11*:2287-2302 (1983)). More preferably, the EGF motif repeat protein of the present invention is expressed as a mature protein (Hitzeman *et al.*, *Nature 293*:717-722 (1981); Valenzuela *et al.*, *Nature 298*:347-350 (1982); and Derynck *et al.*, *Nucleic Acids Res. 11*:1819-1837 (1983)).

Native and engineered yeast promoters suitable for use in the present invention have been reviewed by Romanos *et al.*, *Yeast 8*:423-488 (1992). Most preferably, the EGF motif of the present invention is secreted by the yeast cell (Blobel and Dobberstein, *J. Cell Biol. 67*:835-851 (1975); Kurjan and Herskowitz, *Cell 30*:933-943 (1982); Bostian *et al.*, *Cell 36*:741-751 (1984); Rothman and Orci, *Nature 355*:409-415 (1992); Julius *et al.*, *Cell 32*:839-852 (1983); and Julius *et al.*, *Cell 36*:309-318 (1984)).

Where desired, insect cells may be utilized as host cells in the method of the present invention (See, e.g., Luckow, Protein Eng. J. L. Cleland., Wiley-Liss, New York, NY: 183-2180 (1996) and references cited therein). In addition, general methods for expression of foreign genes in insect cells using baculovirus vectors are described in: O'Reilly *et al.*, *Baculovirus Expression Vectors: A Laboratory Manual*. New York, W.H. Freeman and Company (1992), and King and Possee, *The Baculovirus Expression System: A Laboratory Guide*, London, Chapman & Hall. An expression vector is constructed comprising a baculovirus transfer vector, in which a strong baculovirus promoter (such as the polyhedrin promoter) drives transcription of a eukaryotic secretion signal peptide coding region, which is translationally joined to the coding region for the desired protein. For example, the plasmid pVL1393 (obtained from Invitrogen Corp., San Diego, California) can be used. After construction of the vector carrying the gene encoding desired recombinant protein, two micrograms of this DNA is co-transfected with one microgram of baculovirus DNA into cultured *Spodoptera frugiperda* (Sf9) insect cells. Alternatively, recombinant baculoviruses can be created using a baculovirus shuttle vector system (Luckow *et al.*, *J. Virol. 67*: 4566-4579 (1993)), now marketed as the Bac-To-Bac™ Expression System (Life Technologies, Inc. Rockville, MD). Pure recombinant baculoviruses carrying the desired gene are used to infect cells cultured, for example, in serum-free medium such as Excell 401 (JRH Biosciences,

Lenexa, Kansas) or Sf900-II (Life Technologies, Inc.). The recombinant protein secreted into the medium can be recovered by standard biochemical approaches. Supernatants from mammalian or insect cells expressing the recombinant proteins can be first concentrated using any of a number of commercial concentration units. Proteins accumulating within infected cells can be recovered from cell pastes by standard techniques.

Alternatively, mammalian cells can be used to express the nucleic acid molecules of the present invention. There are a variety of methods, known to those with skill in the art, for introducing genetic material into a host cell. A number of vectors, both viral and non-viral have been developed for transferring genes into primary cells. Several reviews describe the application of viral vectors for gene therapy (Robbins *et al.*, *Trends Biotechnol 16*:35-40 (1998)). Suitable viral vectors include, but are not limited to, adenovirus vectors (including replication-deficient recombinant adenovirus (Berkner, *BioTechniques 6*:616-629 (1988); Berkner, *Current Top. Microbiol. Immunol. 158*:39-66 (1992); Brody and Crystal, *Annal. New York Acad. Sci. 716*:90-103 (1994)), retroviral vectors (including replication deficient recombinant retrovirus (Boris-Lawrie and Temin, *Curr. Opin. Genet. Dev. 3*:102-109 (1993); Boris-Lawrie and Temin, *Annal. New York Acad. Sci. 716*:59-71 (1994); Miller, *Current Top. Microbiol. Immunol. 158*:1-24 (1992)), poxvirus vectors, herpesvirus vectors, adeno associated vectors (AAV), alphavirus vectors, lentivirus vectors, and combination vectors. Non-viral based vectors include protein/DNA complexes (Cristiano *et al.*, *Proc. Natl. Acad. Sci. (U.S.A). 90*:2122-2126 (1993); Curiel *et al.*, *Proc. Natl. Acad. Sci. (U.S.A). 88*:8850-8854 (1991); Curiel, *Annal. New York Acad. Sci. 716*:36-58 (1994)), and electroporation and liposome mediated delivery such as cationic liposomes (Farhood *et al.*, *Annal. New York Acad. Sci. 716*:23-35 (1994)).

Preferably, the nucleic acid molecules of the present invention are cloned into a suitable retroviral vector (*see, e.g.*, Dunbar *et al.*, *Blood 85*:3048-3057 (1995); Baum *et al.*, *J. Hematother. 5*: 323-329 (1996); Bregni *et al.*, *Blood 80*:1418-1422 (1992); Boris-Lawrie and Temin, *Curr. Opin. Genet. Dev. 3*:102-109 (1993); Boris-Lawrie and Temin, *Annal. New York Acad. Sci. 716*:59-71 (1994); Miller, *Current Top. Microbiol. Immunol. 158*:1-24 (1992)), adenovirus vector (Berkner,

BioTechniques 6:616-629 (1988); Berkner, *Current Top. Microbiol. Immunol.* 158:39-66 (1992); Brody and Crystal, *Annal. New York Acad. Sci. 716*:90-103 (1994); Baldwin *et al.*, *Gene Ther. 4*:1142-1149 (1997)), RSV, MuSV, SSV, MuLV (Baum *et al.*, *J. Hematother. 5*: 323-329 (1996)), AAV (Chen *et al.*, *Gene Ther.* 5:50-58 (1998); Hallek *et al.*, *Cytokines Mol. Ther. 2*: 69-79 (1996)), AEV, AMV, or CMV (Griffiths *et al.*, *Biochem. J. 241*: 313-324 (1987)).

Adenovirus vectors for the expression of cytokines in the gut have been reported (Macdonald, *Gut 42*:460-461 (1998)). Adenovirus vectors have also been reported to express alpha interferon in hematopoietic stem cells (Ahmed *et al.*, *Leuk. Res. 22*:119-124 (1998)).

Poxvirus vectors include vaccinia and various avianpox (canarypox and fowlpox) or swinepox viruses. Qin and Chatterjee (*Hum. Gene Ther. 7*:1853-1860 (1996)) report the expression of GM-CSF from a vaccinia virus. A vaccinia and a fowlpox virus have been reported to express model tumor antigens (Irvine *et al.*, *J. Natl. Cancer Inst. 89*:1595-1601 (1997)).

Several herpesvirus vectors have been reported for expression of genes neuronal tissues (Oligino *et al.*, *Gene Ther. 5*:491-496 (1998)) or in hematopoietic cells (Dilloo *et al.*, *Blood 89*:119-127 (1997)).

An adeno-associated virus has been reported to express factor IX in a dog model of hemophilia (Monahan *et al.*, *Gene Ther. 5*:40-49 (1998)). It has also been reported that the insulin gene has been expressed by an adeno-associated virus in a diabetic mouse (Sugiyama *et al.*, *Horm. Metab. Res. 29*:599-603 (1997)).

Retroviruses were the first vectors reported to be used in human gene therapy. Retroviruses have been used to express proteins in many tissue types, including the expression of green fluorescent protein in murine hematopoietic cells (Bagley *et al.*, *Transplantation 65*:1233-1240 (1998)) and gene expression in human cord blood stem cells (Conneally *et al.*, *Blood 91*:3487-3493 (1998)).

Lentivirus vectors consist of HIV and related virus vectors. HIV vectors have been reported for use in the treatment of HIV (Corbeau and Wong-Staal, *Virology 243*:268-274 (1998)). These vectors can integrate into the target cell without the target cell going through mitosis as is required by retrovirus vectors (Miyake *et al.*, *Hum. Gene Ther. 9*:467-475 (1998)).

Alphavirus vectors are a family of virus vectors consisting primarily of Sindbis virus, Semliki Forest virus, and VEE virus vectors. A Sindbis vector system has been reported for gene therapy (Hariharan *et al.*, *J. Virol. 72*:950-958 (1998)). Pushko *et al.* (*Virology 239*:389-401 (1997)) have reported expression of heterologous proteins from a Venezuelan equine encephalitis virus vector system that is useful for vaccine delivery.

Vector systems consisting of retrovirus and adenovirus components have been reported (Feng *et al.*, *Nat. Biotechnol. 15*:866-870 (1997)). This system combines the high transient expression of adenovirus with the integration capabilities of retroviruses.

Vectors can be used for delivery of the naked plasmid DNA expressing the nucleic acid molecules of the present invention. Delivery can, for example, be in an aqueous solution by intramuscular injection or by a gene gun approach. The vectors can also be formulated with a variety of liposomes for delivery. Vectors suitable for naked DNA delivery include, but are not limited to, pCMV that is available from Clonetech (Rodriguez *et al.*, *Journal of Virology 72*:5174-5181 (1998)) and pCI, which is available from Promega (Polo *et al.*, *Nature Biotechnology 16*:517-518 (1998)).

Methods and compositions for transforming a bacteria and other microorganisms are known in the art (see, for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., (1980)).

Technology for introduction of DNA into cells is well known to those of skill in the art. Four general methods for delivering a gene into cells have been described: (1) chemical methods (Graham and van der Eb, *Virology 54*:536-539 (1973)); (2) physical methods such as microinjection (Capecchi, *Cell* 22:479-488 (1980)), electroporation (Wong and Neumann, *Biochem. Biophys. Res. Commun. 107*:584-587 (1982); Fromm *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 82*:5824-5828 (1985); U.S. Patent No. 5,384,253, all of which are incorporated in their entirety); and the gene gun (Johnston and Tang, *Methods Cell Biol. 43*:353-365 (1994)); (3) viral vectors (Clapp, *Clin. Perinatol. 20*:155-168 (1993); Lu *et al.*, *J. Exp. Med. 178*:2089-2096 (1993); Eglitis and Anderson, *Biotechniques, 6*:608-614 (1988), all

of which are incorporated in their entirety); and (4) receptor-mediated mechanisms (Curiel *et al., Hum. Gen. Ther. 3*:147-154 (1992), Wagner *et al., Proc. Natl. Acad. Sci. (U.S.A.) 89*:6099-6103 (1992)).

Transformation can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and combinations of these treatments (see for example Potrykus *et al., Mol. Gen. Genet. 205*:193-200 (1986); Lorz *et al., Mol. Gen. Genet. 199*:178 (1985); Fromm *et al., Nature 319*:791 (1986); Uchimiya *et al., Mol. Gen. Genet. 204*:204 (1986); Marcotte *et al., Nature 335*:454-457 (1988)).

Assays for gene expression based on the transient expression of cloned nucleic acid constructs have been developed by introducing the nucleic acid molecules into cells by polyethylene glycol treatment, electroporation, or particle bombardment (Marcotte *et al., Nature 335*: 454-457 (1988); McCarty *et al., Cell 66:* 895-905 (1991); Hattori *et al., Genes Dev. 6:* 609-618 (1992); Goff *et al., EMBO J. 9:* 2517-2522 (1990)). Transient expression systems may be used to functionally dissect gene constructs (*see generally*, Mailga *et al., Methods in Plant Molecular Biology,* Cold Spring Harbor Press (1995)).

CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

A protein of the present invention may exhibit cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, TI 0, B9, B9/1 1, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, TI 165, HT2, CTLL2, TF- 1, Mo7e Mutz-2, AS-E2 and CMK. The

activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for T cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., 1. Immunol. 149:3778-3783, 1992; Bowman et al., 1. Immunol. 152:1756-1761, 1994. Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol I pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin gamma., Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol I pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol I pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol I pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Acad. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 1--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol I pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9--Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol I pp. 6.13. 1, John Wiley and Sons, Toronto. 1991.

AML Proliferation Assay for Bioactive Human Interleukin-3

The factor-dependent cell line AML 193 is obtained from the American Type Culture Collection (ATCC, Rockville, MD). This cell line, established from a patient with acute myelogenous leukemia, is a growth factor-dependent cell line, which displayed enhanced growth in GM-CSF-supplemented medium (Lange *et al., Blood 70*:192 (1987); Valtieri *et al., J. Immunol. 138*:4042 (1987)). The ability of AML 193 cells to proliferate in the presence of human IL-3 has also been reported. (Santoli *et al., J. Immunol. 139*:348 (1987)). A cell line variant is used, AML 193 1.3, which is adapted for long term growth in IL-3 by washing out the growth factors and starving the cytokine dependent AML 193 cells for growth factors for 24 hours. The cells are then replated at $1x10^5$ cells/well in a 24 well plate in media containing 100 U/mL IL-3. It takes approximately 2 months for the cells to grow rapidly in IL-3. These cells are maintained as AML 193 1.3 thereafter by supplementing tissue culture medium with human IL-3.

AML 193 1.3 cells are washed 6 times in cold Hanks balanced salt solution (HBSS, Gibco, Grand Island, NY) by centrifuging cell suspensions at 250 x g for 10 minutes followed by decantation of the supernatant. Pelleted cells are resuspended in HBSS and the procedure is repeated until six wash cycles are completed. Cells washed six times by this procedure are resuspended in tissue culture medium at a density ranging from 2×10^5 viable cells/mL. This medium is prepared by supplementing Iscove's modified Dulbecco's medium (IMDM, Hazelton, Lenexa, KS) with albumin, transferring, lipids and 2-mercaptoethanol. Bovine albumin (Boehringer-Mannheim, Indianapolis, IN) is added at 500 μg/mL; human transferrin (Boehringer-Mannheim, Indianapolis, IN) is added at 100 μg/mL; soybean lipid (Boehringer-Mannheim, Indianapolis, IN) is added at 50 μg/mL; and 2-mercaptoethanol (Sigma, St. Louis, MO) is added at 5×10^{-5} M.

Serial dilutions of human interleukin-3 or EGF motif repeat proteins are made in triplicate series in tissue culture medium supplemented as stated above in 96 well Costar 3596 tissue culture plates. Each well contained 50 μl of medium containing interleukin-3 or EGF motif repeat proteins once serial dilutions are completed. Control wells contained tissue culture medium alone (negative control). AML 193 1.3 cell suspensions prepared as above are added to each well by pipetting

50 µl (2.5 x 10^4 cells) into each well. Tissue culture plates are incubated at 37°C with 5% CO_2 in humidified air for 3 days. On day 3, 0.5 µCi ^{3}H-thymidine (2 Ci/mM, New England Nuclear, Boston, MA) is added in 50 µl of tissue culture medium. Cultures are incubated at 37°C with 5% CO_2 in humidified air for 18-24 hours. Cellular DNA is harvested onto glass filter mats (Pharmacia LKB, Gaithersburg, MD) using a TOMTEC cell harvester (TOMTEC, Orange, CT), which utilized a water wash cycle followed by a 70% ethanol wash cycle. Filter mats are allowed to air dry and then placed into sample bags to which scintillation fluid (Scintiverse II, Fisher Scientific, St. Louis, MO or BetaPlate Scintillation fluid, Pharmacia LKB, Gaithersburg, MD) is added. Beta emissions of samples from individual tissue culture wells are counted in a LKB BetaPlate model 1205 scintillation counter (Pharmacia LKB, Gaithersburg, MD) and data is expressed as counts per minute of ^{3}H-thymidine incorporated into cells from each tissue culture well. Activity of each human interleukin-3 preparation or EGF motif repeat protein preparation is quantitated by measuring cell proliferation (^{3}H-thymidien incorporation) induced by grated concentrations of interleukin-3 or hematopoietic EGF motif. Typically, concentration ranges from 0.05 pM –10^5 pM are quantitated in these assays. Activity is determined by measuring the dose of interleukin-3 or EGF motif repeat protein which provides 50% of maximal proliferation (EC_{50} = 0.5 x (maximum average counts per minute of ^{3}H-thymidine incorporated per well among triplicate cultures of all concentrations of interleukin-3 tested – background proliferation measured by ^{3}H-thymidine incorporation observed in triplicate cultures lacking interleukin-3). This EC_{50} value is also equivalent to 1 unit of bioactivity. Every assay is performed with native interleukin-3 as a reference standard so that relative activity levels could be assigned.

Typically, the EGF motif repeat proteins are tested in a concentration range of 2000 pM to 0.06 pM titrated in serial 2 fold dilutions.

Activity for each sample is determined by the concentration that gave 50% of the maximal response by fitting a four-parameter logistic model to the data. It is observed that the upper plateau (maximal response) for the sample and the standard with which it is compared did not differ. Therefore relative potency calculation for each sample is determined from EC50 estimations for the sample and the standard

as indicated above. AML 193.1.3 cells proliferate in response to hIL-3, hGM-CSF, and hG-CSF. Therefore the following additional assays are performed for some samples to demonstrate that the G-CSF receptor agonist proteins are active. The proliferation assay is performed with the EGF motif repeat protein plus and minus neutralizing monoclonal antibodies.

TF1 c-mpl ligand dependent proliferation assay

The c-mpl ligand proliferative activity can be assayed using a subclone of the pluripotential human cell line TF1 (Kitamura et al., *J. Cell Physiol* 140:323-334. [1989]). TF1 cells are maintained in h-IL3 (100 U/mL). To establish a sub-clone responsive to c-mpl ligand, cells are maintained in passage media containing 10% supernatant from BHK cells transfected with the gene expressing the 1-153 form of c-mpl ligand (pMON26448). Most the cells die, but a subset of cells survive. After dilution cloning, a c-mpl ligand responsive clone is selected, and these cells are split into passage media to a density of 0.3 x 10^6 cells/mL the day prior to assay set-up. Passage media for these cells is the following: RPMI 1640 (Gibco), 10% FBS (Harlan, Lot #91206), 10% c-mpl ligand supernatant from transfected BHK cells, 1 mM sodium pyruvate (Gibco), 2 mM glutamine (Gibco), and 100 µg/mL penicillin-streptomycin (Gibco). The next day, cells are harvested and washed twice in RPMI or IMDM media with a final wash in the ATL, or assay media. ATL medium consists of the following: IMDM (Gibco), 500 µg/mL of bovine serum albumin, 100 ug/mL of human transferrin, 50 ug/mL soybean lipids, 4 x 10^{-8} M beta-mercaptoethanol and 2 mL of A9909 (sigma, antibiotic solution) per 1000 mL of ATL. Cells are diluted in assay media to a final density of 0.25 x 10^6 cells/mL in a 96-well low evaporation plate (Costar) in a final volume of 50 ul. Transient supernatants (conditioned media) from transfected clones are added at a volume of 50 ul as duplicate samples at a final concentration of 50% and diluted three-fold to a final dilution of 1.8%. Triplicate samples of a dose curve of IL-3 variant pMON13288 starting at 1 ng/mL and diluted using three-fold dilutions to 0.0014ng/mL is included as a positive control. Plates are incubated at 5% CO_2 and 37°C. At day six of culture, the plate is pulsed with 0.5 µCi of 3µl/well (NEN) in a

volume of 20 µl/well and allowed to incubate at 5% CO_2 and 37°C for four hours. The plate is harvested and counted on a Betaplate counter.

MUTZ-2 cell line proliferation assay

MUTZ-2 cells are seeded at $2.5x10^4$ cells per well in microwell plates (Costar) with or without cytokines in serum-free IMDM containing bovine serum albumin (500 mg/ml), human transferrin (100 mg/ml), soybean lipids (50 mg/ml; Boehringer Manheim) and 2-mercaptoethanol (50 mM). After 60 hours, cells are incubated with [*methyl*-^{3}H] thymidine (New England Nuclear) at 0.5 mCi (18.5 kBq) per well for 6 hours and then can be harvested onto a glass fiber filter mat for measurement of radioactivity with a beta counter (Pharmacia LKB).

Other in vitro cell based proliferation assays

Other *in vitro* cell based assays, known to those skilled in the art, may also be useful to determine the activity of the EGF motif repeat protein depending on the factors that comprise the molecule in a similar manner as described in the AML 193.1.3 cell proliferation assay. The following are examples of other useful assays.

TF1 proliferation assay: TF1 is a pluripotential human cell line (Kitamura *et al., J. Cell Physiol 140*:323-334 (1989)) that responds to hIL-3.

32D proliferation assay: 32D is a murine IL-3 dependent cell line which does not respond to human IL-3 but does respond to human G-CSF which is not species restricted.

Baf/3 proliferation assay: Baf/3 is a murine IL-3 dependent cell line which does not respond to human IL-3 or human c-mpl ligand but does respond to human G-CSF which is not species restricted.

T1165 proliferation assay: T1165 cells are a IL-6 dependent murine cell line (Nordan *et al., Science 233*: 566-569 (1986)), which respond to IL-6 and IL-11.

Human Plasma Clot meg-CSF Assay: Used to assay megakaryocyte colony formation activity (Mazur *et al., J. Clin. Invest. 68*: 733-741 (1981)).

Transfected cell lines

Cell lines such as the murine Baf/3 cell line can be transfected with a colony stimulating factor receptor, such as the human G-CSF receptor or human c-mpl receptor, which the cell line does not have. These transfected cell lines can be used to determine the activity of the ligand for which the receptor has been transfected into the cell line.

One such transfected Baf/3 cell line was made by cloning the cDNA encoding c-mpl from a library made from a c-mpl responsive cell line and cloned into the multiple cloning site of the plasmid pcDNA3 (Invitrogen, San Diego, Ca.). Baf/3 cells are transfected with the plasmid via electroporation. The cells are grown under G418 selection in the presence of mouse IL-3 in Wehi-conditioned media. Clones are established through limiting dilution.

In a similar manner the human G-CSF receptor can be transfected into the Baf/3 cell line and used to determine the bioactivity of the hematopoietic EGF motif.

Analysis of c-mpl ligand proliferative activity

1. Bone marrow proliferation assay

a. $CD34^+$ Cell Purification

Bone marrow aspirates (15-20 mL) are obtained from normal allogeneic marrow donors after informed consent. Cells are diluted 1:3 in phosphate buffered saline (PBS, Gibco-BRL), 30 mL are layered over 15 mL Histopaque-1077 (Sigma) and centrifuged for 30 minutes at 300 xs. The mononuclear interface layer is collected and washed in PBS. $CD34^+$ cells are enriched from the mononuclear cell preparation using an affinity column per manufacturer's instructions (CellPro, Inc, Bothell WA). After enrichment, the purity of $CD34^+$ cells is 70% on average as determined by using flow cytometric analysis using anti-CD34 monoclonal antibody conjugated to fluorescein and anti-CD38 conjugated to phycoerythrin (Becton Dickinson, San Jose CA).

Cells are resuspended at 40,000 cells/mL in X-Vivo 10 media (Bio-Whittaker, Walkersville, MD) and 1 mL is plated in 12-well tissue culture plates (Costar). The growth factor rhIL-3 is added at 100 ng/mL (pMON5873) into some wells. HIL3 variants are used at 10 ng/mL to 100 ng/mL. Conditioned media from BHK cells transfected with plasmid encoding c-mpl ligand or EGF motif repeat protein are tested by addition of 100 µl of supernatant added to 1 mL cultures (approximately a 10% dilution). Cells are incubated at 37°C for 8-14 days at 5% CO_2 in a 37°C humidified incubator.

b. Cell Harvest and Analysis:

At the end of the culture period a total cell count is obtained for each condition. For fluorescence analysis and ploidy determination cells were washed in megakaryocyte buffer (MK buffer, 13.6 mM sodium citrate, 1 mM theophylline, 2.2 µm PGE1, 11 mM glucose, 3% w/v BSA, in PBS pH 7.4,) (Tomer *et al.*, *Blood 70*:1735-1742 (1987)) resuspended in 500 µl of MK buffer containing anti-CD41a FITC antibody (1:200, AMC, Westbrook, ME) and washed in MK buffer. For DNA analysis cells are permeablized in MK buffer containing 0.5% TWEEN 20 (Fisher, Fair Lawn, NJ) for 20 min. on ice followed by fixation in 0.5% Tween-20 and 1% paraformaldehyde (Fisher Chemical) for 30 minutes followed by incubation in propidium iodide (Calbiochem, La Jolla Ca) (50 µg/mL) with RNase (400 U/mL) in 55% v/v MK buffer (200 mOsm) for 1-2 hours on ice. Cells are analyzed on a FACScan or Vantage flow cytometer (Becton Dickinson, San Jose, CA). Green fluorescence (CD41a-FITC) is collected along with linear and log signals for red fluorescence (PI) to determine DNA ploidy. All cells were collected to determine the percent of cells that were CD41+. Data analysis is performed using software by LYSIS (Becton Dickinson, San Jose, CA). Percent of cells expressing the CD41 antigen is obtained from flow cytometry analysis (Percent). Absolute (Abs) number of $CD41^+$ cells/mL is calculated by: (Abs) = (Cell Count) * (Percent)/100.

2. Megakaryocyte fibrin clot assay.

$CD34^+$ enriched population is isolated as described above. Cells are suspended at 25,000 cells/mL with or without cytokine(s) in a media consisting of a base Iscoves IMDM media supplemented with 0.3% BSA, 0.4mg/mL apo-transferring, 6.67 μM $FeCl_2$, 25 μg/mL L-asparagine, 500 μg/mL α-amino-n-caproic acid and penicillin/streptomycin. Prior to plating into 35 mm plates, thrombin is added (0.25 Units/mL) to initiate clot formation. Cells are incubated at 37°C for 13 days at 5% CO_2 in a 37°C humidified incubator.

At the end of the culture period plates are fixed with methanol:acetone (1:3), air dried and stored at –20°C until staining. A peroxidase immunocytochemistry staining procedure is used (Zymed, Histostain-SP. San Francisco, CA) employing a cocktail of primary monoclonal antibodies consisting of anti-CD41a, CD42 and CD61. Colonies are counted after staining and classified as negative, CFU-MK (small colonies, 1-2 foci and less that approx. 25 cells), BFU-MK (large, multi-foci colonies with > 25 cells) or mixed colonies (mixture of both positive and negative cells.

Methylcellulose Assay

This assay reflects the ability of colony stimulating factors to stimulate normal bone marrow cells to produce different types of hematopoietic colonies *in vitro* (Bradley *et al.*, *Aust. Exp. Biol. Sci. 44*:287-300 (1966), Pluznik *et al.*, *J. Cell Comp. Physiol 66*:319-324 (1965)).

Methods

Approximately 30 mL of fresh, normal, healthy bone marrow aspirate is obtained from individuals following informed consent. Under sterile conditions samples are diluted 1:5 with a 1X PBS (#14040.059 Life Technologies, Gaithersburg, MD.) solution in a 50 mL conical tube (#25339-50 Corning, Corning MD). Ficoll (Histopaque 1077 Sigma H-8889) is layered under the diluted sample and centrifuged, 300 Xg for 30 min. The mononuclear cell band is removed and washed two times in 1X PBS and once with 1% BSA PBS (CellPro Co., Bothel, Wa). Mononuclear cells are counted and $CD34^+$ cells are selected using the Ceprate

LC (CD34) Kit (CellPro Co., Bothel, WA) column. This fractionation is performed since all stem and progenitor cells within the bone marrow display CD34 surface antigen.

Cultures are set up in triplicate with a final volume of 1.0 mL in a 35 x 10 mm Petri dish (Nunc#174926). Culture medium is purchased from Terry Fox Labs. (HCC-4230 medium (Terry Fox Labs, Vancouver, B.C., Canada) and erythropoietin (Amgen, Thousand Oaks, CA.) is added to the culture media. 3,000-10,000 $CD34^+$ cells are added per dish. Recombinant IL-3, purified from mammalian cells for *E. coli*, and hematopoietic growth factor proteins, in conditioned media from transfected mammalian cells or purified from conditioned media from transfected mammalian cells or *E. coli*, are added to give final concentrations ranging from 0.001 nM to 10 nM. Recombinant hIL-3, GM-CSF, c-mpl ligand, and EGF motif repeat proteins are supplied Monsanto. G-CSF (Neupogen) is from Amgen (Thousand Oaks Calf.). Cultures are resuspended using a 3-cc syringe and 1.0 mL is dispensed per dish. Control (baseline response) cultures received no colony stimulating factors. Positive control cultures receive conditioned media (PHA stimulated human cells: Terry Fox Lab. H2400). Cultures are incubated at 37°C, 5% CO_2 in humidified air. Hematopoietic colonies that are defined as greater than 50 cells are counted on the day of peak response (days 10-11) using a Nikon inverted phase microscope with a 40x objective combination. Groups of cells containing fewer than 50 cells are referred to as clusters. Alternatively colonies can be identified by spreading the colonies on a slide and stained or they can be picked, resuspended, and spun onto cytospin slides for staining.

AS-E2 Cell Proliferation Assay

The factor-dependent cell line AS-E2 cell line, established from a patient with acute myeloid leukemia, is a growth factor dependent cell line which displayed enhanced growth in erythropoietin (EPO)-supplemented media (Miyazaki, Y., Kuriyama, K., Higuchi, M., Tsushima, H., Sohda, H., Imai, N., Saito, M., Kondo, T., and Tomonaga, M. Establishment and characterization of a new erythropoietin-dependent acute myeloid leukemia cell line, AS-E2. *Leukemia*, 11;1941-9, 1997).

This cell line also demonstrated enhanced proliferation in response to HS-5 CM and was used to measure HS-5 induced growth.

AS-E2 cells were maintained at a density of 1 x 10^5 to 5 x 10^5 cells/ml in tissue culture medium prepared by IMDM with 20% fetal bovine serum (FBS, Harlan, Indianapolis, IN), 4 ml/liter of 2-mercaptoethanol (Sigma, St. Louis, MO), and 1 unit/ml EPO (Epogen, Amgen, Thousand Oaks, CA or Procrit, Ortho Biotech, Raritan, NJ). Every 3 months the AS-E2 cell cultures were replaced by new cultures initiated from frozen stocks.

For cell proliferation assays, AS-E2 cells in culture media were harvested by centrifuging the cell suspensions at 250 X g for 5 minutes. Cells were washed twice in cold Dulbecco's phosphate-buffered saline by resuspension and centrifugation as above. The final cell pellet was resuspended in assay media at a density for 1 x 10^4 to 5 x 10^4 viable cells/50 ml. The assay medium was prepared by supplementing Iscove's modified Dulbecco's Medium (IMDM, Gibco/BRL, Grand Island, NY) with 20% fetal bovine serum (FBS, Harlan, Indianapolis, IN), 4 ml/liter of 2-mercaptoethanol (Sigma, St. Louis, MO), and 1 ml/liter of penicillin/ streptomycin solution (10,000 units penicillin/ml and 10 mg streptomycin per ml of normal saline, Sigma, St. Louis, MO).

Human Cord Blood Hematopoietic Growth Factor Assays

Bone marrow cells are traditionally used for *in vitro* assays of hematopoietic colony stimulating factor (CSF) activity. However, human bone marrow is not always available, and there is considerable variability between donors. Umbilical cord blood is comparable to bone marrow as a source of hematopoietic stem cells and progenitors (Broxmeyer *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 89*:4109-113 (1992); Mayani *et al.*, *Blood 81*:3242-3258 (1993)). In contrast to bone marrow, cord blood is more readily available on a regular basis. There is also a potential to reduce assay variability by pooling cells obtained fresh from several donors, or to create a bank of cryopreserved cells for this purpose. By modifying the culture conditions, and/or analyzing for lineage specific markers, it is possible to assay

specifically for granulocyte/macrophage colonies (CFU-GM), for megakaryocyte CSF activity, or for high proliferative potential colony forming cell (HPP-CFC) activity.

Methods

Mononuclear cells (MNC) are isolated from cord blood within 24 hr. of collection, using a standard density gradient (1.077 g/mL Histopaque). Cord blood MNC is further enriched for stem cells and progenitors by several procedures, including immunomagnetic selection for $CD14^-CD34^+$ cells; panning for SBA-, $CD34^+$ fraction using coated flasks from Applied Immune Science (Santa Clara, CA); and $CD34^+$ selection using a CellPro (Bothell, WA) avidin column. Either freshly isolated or cryopreserved $CD34^+$ cell enriched fractions are used for the assay. Duplicate cultures for each serial dilution of sample (concentration range from 1 pM to 1204 pM) are prepared with $1x10^4$ cells in 1 ml of 0.9% methycellulose containing medium without additional growth factors (Methocult H4230 from Stem Cell Technologies, Vancouver, BC.). In some experiments, Methocult H4330 containing erythropoietin (EPO) is used instead of Methocult H4230, or Stem Cell Factor (SCF), 50 ng/mL (Biosource International, Camarillo, CA) is added. After culturing for 7-9 days, colonies containing >30 cells are counted. In order to rule out subjective bias in scoring, assays are scored blind.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095,1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

IMMUNE STIMULATING OR SUPPRESSING ACTIVITY

A protein of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by vital (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, *Leishmania* spp., *malaria* spp. and various fungal infections such as *Candidiasis*. Of course, in this regard, a protein of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer. Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft- versus-host disease, and autoimmune inflammatory eye disease. Such a protein of the present invention may also to be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein of the present invention.

Using the proteins of the invention it may also be possible to regulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance

in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process, which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and is organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (e.g., B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding co-stimulatory signal. Blocking B lymphocyte antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of co-stimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have

been used to examine the immunosuppressive effects of CTLA41g fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function in vivo on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self-tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block co-stimulation of T cells by disrupting receptor:ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen- specific tolerance of autoreactive T cells, which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory forms of B lymphocyte antigens systemically.

Alternatively, anti-vital immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APC's either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I α chain protein and β sub.2 microglobulin protein or a MHC class II α chain protein and a MHC class II β chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7- 1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley- Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564- 1572, 1985; Takai et al., 1. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 198 1; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., J. Immunol. 137:3494-3500,1986; Bowmanet al., J. Virology 61:1992-1998; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., Cellular Immunology 133:327-341, 199 1; Brown et al., J. Immunol. 153:3 079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Thl and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley- Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 13 7:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed 30 by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba

et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260,1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264,1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648,1992.

Assays for proteins that influence early steps of T-cell commitment and development 15 include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

HEMATOPOIESIS REGULATING ACTIVITY

A protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting

the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous) as normal cells or genetically manipulated for gene therapy.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In Culture of Hematopoietic Cells. R. 1. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In Culture of Hematopoietic Cells. R. 1. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc.,

New York, N.Y. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In Culture of Hematopoietic Cells. R. 1. Freshney, et al. eds. Vol pp. 1-2 1, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R. 1. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In Culture of Hematopoietic Cells. R. 1. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

TISSUE GROWTH ACTIVITY

A protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of bums, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells, or induce differentiation of progenitors of bone- forming cells. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation

or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors ex vivo for return in vivo to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions that

may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma, and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non- healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to regenerate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. W095/16035 (bone, cartilage, tendon); International Patent Publication No. W095/05846 (nerve, neuronal); International Patent Publication No. W091/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. 1. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

ACTIVIN/INHIBIN ACTIVITY

A protein of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a protein of the present invention, alone or in heterodimers with a member of the inhibin α-family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-D group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep, and pigs.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A protein of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic proteins can be used to

mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes, or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates 3.0 and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12. 1 - 6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A protein of the invention may also exhibit hemostatic or thrombolytic activity. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Such a protein is expected to be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance

coagulation and other hemostatic events in treating wounds resulting from trauma, surgery, or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assay for hemostatic and thrombolytic activities include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1996; Burdick et al., Thrombosis Res. 45:413 -419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474,1988.

RECEPTOR/LIGAND ACTIVITY

A protein of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selecting, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing

Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1-7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

ANTI-INFLAMMATORY ACTIVITY

Proteins of the present invention may also exhibit anti-inflammatory activity. The anti- inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL- 1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

LEUKEMIAS

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic,

monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions, which may be treated in a patient (including human and non-human mammalian patients) according to the invention include but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions that sever a portion of the nervous system, or compression injuries;

(ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system 5results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

(iii) malignant lesions, in which a portion of the nervous system is destroyed or injured by malignant tissue that is either a nervous system associated malignancy or a malignancy derived from non-nervous system tissue;

(iv) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

(v) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to

degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

(vi) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B 12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

(vii) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

(viii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

(ix) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

(i) increased survival time of neurons in culture;

(ii) increased sprouting of neurons in culture or in vivo;

(iii) increased production of a neuron-associated molecule in culture or in vivo, e.g., 10 choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or

(iv) decreased symptoms of neuron dysfunction in vivo.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased

sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, etc., depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, e.g., weakness, motor neuron conduction velocity, or functional disability.

In a specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot- Marie-Tooth Disease).

OTHER ACTIVITIES

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or caricadic cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress,

cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain is reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

A protein of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may also contain (in addition to protein and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. The pharmaceutical composition may further contain other agents, which either enhance the activity of the protein or compliment its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein of the invention, or to minimize side effects. Conversely, protein of the present invention

may be included in formulations of the particular cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent to minimize side effects of the cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent. A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, e.g., treatment, healing, prevention, or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein of the present invention is administered to a mammal having a condition to be treated. Protein of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines, or other hematopoietic factors. When co- administered with one or more cytokines, lymphokines or other hematopoietic factors, protein of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein of the present invention

in combination with cytokine(s), lymphokine(s), and other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries, which facilitate processing of the active compounds into preparations which, can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, e.g., by means of conventional mixing, dissolving,

granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein of the present invention is administered orally, protein of the present invention will be in the form of a tablet, capsule, powder, solution, or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein of the present invention, and preferably from about 25 to 90% protein of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other Saccharide solution, or glycols such as ethylene glycol, propylene glycol, or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein of the present invention, and preferably from about 1 to 50% protein of the present invention.

When a therapeutically effective amount of protein of the present invention is administered by intravenous, cutaneous, or subcutaneous injection, protein of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skilled in the art. For injection, the agents of the invention may be formulated in 5aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate

to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic arid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations, which can be used orally, include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the

compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide, or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, e.g., gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, e.g., in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions, or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing, and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances, which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, e.g., containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such

long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a cosolvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The cosolvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD: 5W) consists of VPD diluted 1: 1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co- solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, e.g. polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the proteinase inhibiting compounds of the invention may be provided as salts with pharmaceutically compatible counterions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention. The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithin, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S.

Pat. Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

The amount of EGF motif repeat protein in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments that the patient has undergone. Ultimately, the attending physician will decide the amount of protein of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein of the present invention and observe the patient's response. Larger doses of protein of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 μg to about 100 mg (preferably about 0. 1μg to about 10 mg, more preferably about 0. 1 μg to about 1 mg) of protein of the present invention per kg body weight. For compositions of the present invention, which are useful for bone, cartilage, tendon or ligament regeneration, and the therapeutic method, includes administering the composition topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage, or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein of the invention, which may also optionally be included in the composition, as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance, and interface properties. The particular

application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, tricalciumphosphate, hydroxyapatite, polylactic acid, polyglycolic acid, and polyanhydrides. Other potential matrices are biodegradable and biologically well defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are non-biodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above-mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalciumphosphate. The bioceramics may be altered in composition, such as in calcium- aluminate-phosphate and processing to alter pore size, particle sizes, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl is cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer, and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0. 5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorbtion of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as

epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF-.alpha. and TGF-.beta.), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, e.g., amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (e.g., bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF 1 (insulin like growth factor 1), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations, and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either in vivo or ex vivo into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured ex vivo in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced in vivo for therapeutic purposes.

EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically

effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amounts is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (ie., the concentration of the test compound, which achieves a half-maximal inhibition of the C- proteinase activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50}. Compounds that exhibit high therapeutic indices are preferred. The data obtained from these cell cultures assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. See, e.g., Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p. 1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety, which are sufficient to maintain the C-proteinase inhibiting effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from in vitro data; for example, the concentration necessary to achieve 50-90% inhibition of the C-proteinase using the assays described herein. Dosages necessary to achieve the MEC will depend on individual characteristics and route of

administration. However, HPLC assays or bioassays can be used to determine plasma concentrations. Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen, which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90%, and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration. The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device, which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

ANTI BODIES

Another aspect of the invention is an antibody that specifically binds the polypeptide of the invention. Such antibodies can be either monoclonal or polygonal antibodies, as well fragments thereof, and humanized forms or fully human forms, such as those produced in transgenic animals. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention. Protein of the invention may also be used to immunize animals to obtain polyclonal and monoclonal antibodies, which specifically react with the protein. Such antibodies may be obtained using either the entire protein or fragments thereof as an immunogen. The peptide immunogens additionally may contain a cysteine

residue at the carboxyl terminus, and are conjugated to a hapten such as keyhole limpet hemocyanin (KLH). Methods for synthesizing such peptides are known in the art, for example, as in R. P. Merrifield, J. Amer. Chem. Soc. 85, 2149-2154 (1963); J. L. Krstenansky, et al., FEBS Lett. 211, 10 (1987). Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein. In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A.M., Monoclonal Antibodies Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984): St. Groth et al., J Immunol 35:1-21 (1990); Kohler and Milstein, Nature 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor et al, Immunology Today 4:72 (1983); Cole et al., in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. (1985), pp. 77-96).

Any animal (mouse, rabbit, etc.), which is known to produce antibodies, can be immunized with a peptide or polypeptide of the invention. Methods for immunization are well known in the art. Such methods include subcutaneous or intraperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal, which is immunized, the antigenicity of the peptide and the site of injection. The protein that is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to, coupling the antigen with a heterologous

protein (such as globulin or 0-galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Agl4 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells. Any one of a number of methods well known in the art can be used to identify the hybridoma cell, which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz et al., Exp. Cell Research. 1 75:109-124 (1988)). Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A.M., Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)). Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

For polyclonal antibodies, antibody-containing antiserum is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures. The present invention further provides the above-described antibodies in delectably labeled form. Antibodies can be delectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, etc.), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, etc.) fluorescent labels (such as FITC or rhodamine, etc.), paramagnetic atoms, etc. Procedures for accomplishing such labeling are well known in the art, for example, see (Sternberger, L.A. et al., J Histochem. Cytochem. 18:315 (1970); Bayer, E.A. et al., Meth. Enzym. 62:308 (1979); Engval, E. et al., Immunol 109:129 (1972); Goding, J.W. J Immunol Meth. 13:215 (1976)).

The labeled antibodies of the present invention can be used for *in vitro, in vivo,* and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid

supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D.M. et al., "Handbook of Experimental Immunology " 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro, in vivo,* and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

Another aspect of the present invention is the use of an EGF motif repeat protein in both *ex vivo* and *in vivo* proliferation and/or expansion of stem cells. Several methods for *ex vivo* expansion of stem cells have been reported. Such selection methods and expansion methods use various colony stimulating factors including c-kit ligand (Brandt *et al.*, *Blood 83*:1507-1514 (1994); McKenna *et al.*, *Blood 86*:3413-3420 (1995)), IL-3 (Brandt *et al.*, *Blood 83*:1507-1514 (1994); Sato *et al.*, *Blood 82*:3600-3609 (1993)), G-CSF (Sato *et al.*, *Blood 82*:3600-3609 (1993)), IL-1 (Muench *et al.*, *Blood 81*:3463-3473 (1993)), IL-6 (Sato *et al.*, *Blood 82*:3600-3609 (1993)), IL-11 (Lemoli *et al.*, *Exp. Hem. 21*:1668-1672 (1993); Sato *et al.*, *Blood 82*:3600-3609 (1993)), flt-3 ligand (McKenna *et al.*, *Blood 86*:3413-3420 (1995)), GM-CSF (Haylock et al., *Blood 80*:1405-1412 (1992); Sato et al., *Blood 82*:3600-3609 (1993)); EPO (Fisher, *P.S.E.B.M. 216*:358-366 (1997); TPO (Schipper et. al., *Br. J. Haematol. 101*:425-435 (1998); and/or combinations thereof (Brandt *et al.*, *Blood 83*:1507-1514 (1994); Haylock *et al.*, *Blood 80*:1405-1412 (1992); Koller *et al.*, *Biotechnology 11*:358-363 (1993); Lemoli *et al.*, *Exp. Hem. 21*:1668-1672 (1993); McKenna *et al.*, *Blood 86*:3413-3420 (1995); Muench *et al.*, *Blood 81*:3463-3473 (1993); Patchen *et al.*, *Biotherapy 7*:13-26 (1994); Sato *et al.*, *Blood 82*:3600-3609 (1993); Smith *et al.*, *Exp. Hem. 21*:870-877 (1993); Steen *et al.*, *Stem Cells 12*:214-224 (1994); Tsujino *et al.*, *Exp. Hem. 21*:1379-1386 (1993)). Although hIL-3 has been reported to be the most potent growth factor in expanding peripheral blood $CD34^+$ cells (Sato *et al.*, *Blood 82*:3600-3609 (1993); Kobayashi *et al.*, *Blood 73*:1836-1841 (1989)). No single factor has been reported to be as effective as a combination of multiple factors. Therefore, under an embodiment of the present invention, the EGF motif repeat protein of the present invention is co-

administered with at least one other hematopoietic growth factor, cytokine, or interleukin.

Under one embodiment, the nucleic acid molecules of the present invention are used as a surrogate marker to measure the activity of another growth factor. All growth factors are capable of both *ex vivo* and *in vivo* proliferation and/or expansion of stem cells. Therefore, the nucleic acid molecules of the present invention are used, as a surrogate marker, to measure the proliferation and/or expansion activity of a growth factor.

It is understood that the expressed EGF motif repeat protein of the present invention may be used for the *ex vivo* expansion of hematopoietic progenitor cells.

Bone marrow transplants have been used to treat patients with neutropenia and thrombocytopenia. Several problems associated with the use of bone marrow in reconstitution of a compromised hematopoietic system include: the limited number of stem cells in bone marrow, spleen, or peripheral; Graft Versus Host Disease; graft rejection; and possible contamination with tumor cells. It has been reported that stem cells exhibit a dose response such that the greater the number of cells, the more enhanced hematopoietic recovery. Therefore, the *ex vivo* expansion of stem cells should enhance hematopoietic recovery and thereby enhance patient survival.

Graft Versus Host disease and graft rejection have been reported for even HLA- matched sibling donors. An alternative to allogenic none marrow transplants is autologous bone marrow transplants. In autologous bone marrow transplants, a portion of a patient's marrow is harvested prior to myeloblative therapy, e.g., high-dose chemotherapy and radiation therapy, and is transplanted back into the patient after completion of the myeloblative therapy regimen. Autologous transplantation eliminates the risk of Graft Versus Host Disease and graft rejection. However, autologous bone marrow transplants still present problems in terms of the limited number of stem cells in the marrow and the potential contamination with tumor cells. The limited number of stem cells may be overcome by *ex vivo* expansion of the stem cells.

Under another embodiment of the present invention the *ex vivo* expanded hematopoietic progenitor cells are used in bone marrow transplantation (Kessinger and Armitage, *Blood 77*:211-213 (1991)). By expanding the hematopoietic

progenitor cells, it is possible to reduce the number and duration of leucapheresis procedures required during autologous transplantation, thereby reducing the risk of disease contamination in the apheresis products (Alcorn and Holyoake, *Blood Rev. 10*:167-176 (1996)).

Under another embodiment of the present invention, the EGF motif repeat protein of the present invention is used for the *ex vivo* expansion of umbilical cord blood cells (Lu *et al.*, *Exp. Hematol. 21*:1442-1446 (1993); Westwood *et al.*, *Br. J. Haematol. 86*:468-474 (1994); Traycoff *et al.*, *Blood 85*:2059-2068 (1995); Reems *et al.*, *Bio. Blood Marrow Transp. 3*:133-141 (1997)). The expanded umbilical cord blood cells can then be used in allogenic transplantation. Alternatively, the EGF motif repeat protein of the present invention is used for the *ex vivo* expansion of fetal tissue. Reems *et al.* describes the *ex vivo* expansion of cord blood cells in the presence of an HS-5 supernatant and kit ligand (Reems *et al.*, *Bio. Blood Marrow Transp. 3*:133-141 (1997)).

Under one embodiment, the *ex vivo* expansion of the hematopoietic progenitor cells occurs in a stromal or stroma-free long term culture in the presence of various combinations of interleukins, granulocyte macrophage colony stimulating factor and the EGF motif repeat protein of the present invention. A stroma-free, cytokine-based culture is preferable, in that cultures can be established under relatively defined serum-free conditions and cell proliferation and differentiation can be manipulated according to the hematopoietic growth factor(s) employed (Mayani *et al.*, *Blood 82*:2664-2672 (1994); Rebel *et al.*, *Blood 83*:128-136 (1994); Ploemacher *et al.*, *Leukemia 7*:1381-1388 (1993); Henschler *et al.*, *Blood 84*:2898-2903 (1994)). The *ex vivo* expansion of the hematopoietic progenitor cells in a stroma-free long term culture has been described by De Bruyn *et al.*, *J. Hematotherapy 6*:93-102 (1997). The *ex vivo* expansion of the hematopoietic progenitor cells in a stromal long term culture has been described by Dexter *et al.*, *J. Cell Physiol 91:335-344* (1977). More preferably, the *ex vivo* method for expansion of the hematopoietic progenitor cells is capable of sustaining long term *ex vivo* expansion.

Colony stimulating factors (CSFs), such as hIL-3, have been administered alone, co-administered with other CSFs, or in combination with bone marrow

transplants subsequent to high dose chemotherapy to treat neutropenia and thrombocytopenia, which are often the result of such treatment (Bacigalupo, *Eur. J. Cancer 30*:S26-S29 (1994)). The administration of CSFs, however, has not been sufficient to completely eliminate neutropenia and thrombocytopenia. The myeloid lineage, which is comprised of monocytes (macrophages), granulocytes (including neutrophils) and megakaryocytes, plays a role in preventing potentially life-threatening infections and bleeding.

The EGF motif repeat protein of the present invention may be used in the treatment of diseases characterized by decreased levels of either myeloid, erythroid, lymphoid, or megakaryocyte cells, or a combination thereof, within the hematopoietic system. In addition, they may be used to activate mature myeloid and/or lymphoid cells. Conditions or diseases that can be treated or ameliorated by the EGF motif repeat protein of the present invention include, but are not limited to, leukopenia, neutropenia, aplastic anemia, cyclic neutropenia, idiopathic neutropenia, Chediak-Higashi syndrome, systemic lupus erythematosus (SLE), leukemia, myelodysplastic syndrome, myelofibrosis, and thrombocytopenia.

Leukopenia is a reduction in the number of circulating leukocytes (white blood cells) that has been reported to be induced by exposure to certain viruses or radiation. It is often a reported side effect of various forms of cancer therapy, e.g., exposure to chemotherapeutic drugs or radiation treatments, and of infection or hemorrhage. Therapeutic treatment of leukopenia with the EGF motif repeat protein of the present invention may avoid or ameliorate undesirable side effects caused by treatment with the various forms of cancer therapy.

The only reported therapy for thrombocytopenia is platelet transfusions, which are costly and carry the significant risks of infection (e.g., HIV, HBV, etc.) and alloimmunization. The EGF motif repeat protein of the present invention may alleviate or diminish the need for platelet transfusion.

The EGF motif repeat protein of the present invention may be used in the mobilization of hematopoietic progenitor and stem cells in peripheral blood. Peripheral blood derived progenitors have been reported to be effective in reconstituting patients who have undergone autologous bone marrow transplantation. The hematopoietic growth factors G-CSF and GM-CSF have been

reported to enhance the number of circulating progenitor and stem cells in the peripheral blood. This has simplified the procedure for peripheral stem cell collection and decreased the cost of the procedure by decreasing the number of pheresis required. The EGF motif repeat protein of the present invention may be used in mobilizing stem cells and further enhancing the efficiency of peripheral stem cell transplantation.

One aspect of the present invention provides a method for selective *ex vivo* expansion of stem cells. The term "stem cell" refers to totipotent hematopoietic stem cells, as well as precursors and progenitor cells, which can be isolated from bone marrow, spleen, or peripheral blood. The term "expansion" refers to the differentiation and proliferation of the cells. The present invention provides a method for selective *ex vivo* expansion of stem cells comprising: (a) culturing isolated and purified stem cells with a selective media which contains a EGF motif repeat protein and (b) harvesting said stem cells.

Stem cells, as well as progenitor cells committed to becoming neutrophils, erythrocytes, platelets, etc. may be distinguished from most other cells by the presence or absence of particular progenitor marker antigens, such as CD34, that are present on the surface of these cells and/or by morphological characteristics. The phenotype for a highly enriched human stem cell population is reported as $CD34^+$, $CD38^-$, $Thy\text{-}1^+$, and lin^-. It is to be understood that the present invention is not limited to this stem cell population.

$CD34^+$ enriched human stem cells can be separated by a number of reported methods, including affinity columns or beads, magnetic beads or flow cytometry using antibodies directed to surface antigens such as $CD34^+$. Further, physical separation methods such as counterflow elutriation may be used to enrich hematopoietic progenitor cells. $CD34^+$ progenitors are reported to be heterogeneous, and may be divided into several sub-populations characterized by the presence or absence of co-expression of different lineage associated cell surface associated molecules. Very immature progenitor cells are not reported to express any reported lineage associated markers, such as HLA-DR or CD38, but they may express CD90 (thy-1). Other surface antigens such as CD33, CD38, CD41, CD71, HLA-DR or c-kit can be used to selectively isolate hematopoietic progenitor cells.

Cell surface antigens that are up regulated in, or specifically expressed on, cancer cells have been reported (Greipp and Witzig, *Curr. Opinion Oncol. 8*:20-27 (1996)).

The separated cells can be incubated in selected medium in a culture flask, sterile bag, or in hollow fibers. Various colony stimulating factors may be utilized in order to selectively expand the cells. Factors which have been reported in the *ex vivo* expansion of bone marrow include, but are not limited to, GM-CSF, G-CSF, c-mpl ligand (also known as TPO or MGDF), M-CSF, erythropoietin (EPO), IL-1, IL-4, IL-2, IL-3, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-15, IL-16, LIF, flt3 ligand, stem cell factor (SCF) also known as steel factor or c-kit ligand or combinations thereof.

Another aspect of the invention provides methods of sustaining and/or expanding hematopoietic precursor cells which include inoculating the cells into a culture vessel which contains a culture medium that has been conditioned by exposure to a stromal cell line especially HS-5 (WO 96/02662, Roecklein and Torok-Strob, Blood 85:997-1105, 1995) that has been supplemented with the EGF motif repeat protein of the present invention.

Many drugs have been reported to cause bone marrow suppression or hematopoietic deficiencies. Examples of such drugs are AZT, DDI, alkylating agents and anti-metabolites used in chemotherapy, antibiotics such as chloramphenicol, penicillin, gancyclovir, daunomycin and sulfa drugs, phenothiazones, tranquilizers such as meprobamate, analgesics such as aminopyrine and dipyrone, anti-convulsants such as phenytoin or carbamazepine, antithyroids such as propylthiouracil and methimazole and diuretics. The EGF motif repeat protein of the present invention may be useful in preventing or treating the bone marrow suppression or hematopoietic deficiencies which often occur in patients treated with these drugs.

Hematopoietic deficiencies may also occur as a result of viral, microbial, or parasitic infections and as a result of treatment for renal disease or renal failure, e.g., dialysis. The EGF motif repeat protein of the present invention may be useful in treating such hematopoietic deficiencies.

The treatment of hematopoietic deficiency may include administration of a pharmaceutical composition containing the EGF motif repeat protein to a patient.

The EGF motif repeat protein of the present invention may also be useful for the activation and amplification of hematopoietic precursor cells by treating these cells in vitro with the EGF motif repeat proteins of the present invention prior to injecting the cells into a patient.

Various immunodeficiencies, e.g., in T and/or B lymphocytes, or immune disorders, e.g., rheumatoid arthritis, may also be beneficially affected by treatment with the EGF motif repeat protein of the present invention. Immunodeficiencies may be the result of viral infections, e.g., HTLVI, HTLVII, HTLVIII, severe exposure to radiation, cancer therapy or the result of other medical treatment. The EGF motif repeat protein of the present invention may also be employed, alone or in combination with other colony stimulating factors, in the treatment of other blood cell deficiencies, including thrombocytopenia (platelet deficiency), or anemia. Other uses for these polypeptides are the *in vivo* and *ex vivo* treatment of patients recovering from bone marrow transplants, and in the development of monoclonal and polyclonal antibodies generated by standard methods for diagnostic or therapeutic use.

Other aspects of the present invention are methods and therapeutic compositions for treating the conditions referred to above. Such compositions comprise a therapeutically effective amount of one or more of the EGF motif repeat protein of the present invention in a mixture with a pharmaceutically acceptable carrier. This composition can be administered either parenterally, intravenously or subcutaneously. When administered, the therapeutic composition for use in this invention is preferably in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such a parenterally acceptable protein solution, having due regard to pH, isotonicity, stability and the like, is within the skill of the art.

The dosage regimen involved in a method for treating the above-described conditions will be determined by the attending physician considering various factors which modify the action of drugs, e.g., the condition, body weight, sex and diet of the patient, the severity of any infection, time of administration and other clinical factors. Generally, a daily regimen may be in the range of 0.2 - 150 μg/kg of EGF motif repeat protein per kilogram of body weight. Preferably, the daily regimen is in

the range of 0.2 - 100 μg/kg of EGF motif repeat protein per kilogram of body weight. More preferably, the daily regimen is in the range of 0.5 - 50 μg/kg and most preferably it is in the range of 0.5 - 10 μg/kg of EGF motif repeat protein per kilogram of body weight. Dosages would be adjusted relative to the activity of a given EGF motif repeat protein and it would not be unreasonable to note that dosage regimens may include doses as low as 0.1 microgram and as high as 1 milligram per kilogram of body weight per day. In addition, there may exist specific circumstances where dosages of EGF motif repeat protein would be adjusted higher or lower than the range of 0.2 - 150 micrograms per kilogram of body weight. These include co-administration with other colony stimulating factors of IL-3 variants or growth factors; co-administration with chemotherapeutic drugs and/or radiation; the use of glycosylated EGF motif repeat protein; and various patient-related issues mentioned earlier in this section.

As indicated above, the therapeutic method and compositions may also include co-administration with other human factors. A non-exclusive list of other appropriate colony stimulating factors (CSFs), cytokines, lymphokines, hematopoietic growth factors and interleukins for simultaneous or serial co-administration with the polypeptides of the present invention includes GM-CSF, G-CSF, c-mpl ligand (also known as TPO or MGDF), M-CSF, erythropoietin (EPO), IL-1, IL-4, IL-2, IL-3, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-15, IL-16, LIF, flt3 ligand, stem cell factor (SCF) also known as steel factor or c-kit ligand, or combinations thereof. The dosage recited above would be adjusted to compensate for such additional components in the therapeutic composition. Progress of the treated patient can be monitored by periodic assessment of the hematological profile, e.g., differential cell count and the like.

Under one embodiment, the nucleic acid molecules of the present invention are used as a surrogate marker to measure the effectiveness of a therapeutic composition for treating hematopoietic deficiencies. Alternatively, the nucleic acid molecules of the present invention can be used as a surrogate marker to measure the effect of a drug at causing hematopoietic deficiencies. Such measurements can be performed, for example, by measuring the presence, absence, or the level of

expression of a nucleic acid molecule complementary to a nucleic acid molecule of the present invention.

EXAMPLES

Having now generally described the invention, the same will be more readily understood through reference to the following examples, which are provided by way of illustration, and are not intended to be limiting of the present invention, unless specified.

Example 1

The HS-5, HS-21, HS-22, HS-23, and HS-27 cells are separately frozen in liquid nitrogen and the mRNA is isolated using known RNA isolation methods. The isolated RNA is stored at -80°C until subsequent manipulation.

The stored RNA is purified using Trizol reagent from Life Technologies (Gaithersburg, Maryland), essentially as recommended by the manufacturer. Poly A+ RNA (mRNA) is purified using magnetic oligo dT beads essentially as recommended by the manufacturer (Dynabeads, Dynal Corporation, Lake Success, New York).

Construction of cDNA libraries is well known in the art and a number of cloning strategies exist. A number of cDNA library construction kits are commercially available. The PCR-Select Differential Screening Kit (CLONTECH Laboratories, Inc., Palo Alto, CA) is used, following the conditions suggested by the manufacturer.

Normalized libraries are made using essentially the Soares procedure (Soares *et. al., Proc. Natl. Acad. Sci. (U.S.A.) 91:* 9228-9232 (1994)). This approach is designed to reduce the initial 10,000-fold variation in individual cDNA frequencies to achieve abundances within one order of magnitude, while maintaining the overall sequence complexity of the library. In the normalization process, the prevalence of high-abundance cDNA clones decreases dramatically, clones with mid-level

abundance are relatively unaffected, and clones for rare transcripts are effectively increased in abundance.

Example 2

Bacteria harboring the cDNA libraries are plated on LB agar containing the appropriate antibiotics for selection and incubated at 37°C for a sufficient time to allow the growth of individual colonies. Single colonies are individually placed in each well of a 96-well microtiter plates containing LB liquid including the selective antibiotics. The plates are incubated overnight at approximately 37°C with gentle shaking to promote growth of the cultures. The plasmid DNA is isolated from each clone using Qiaprep plasmid isolation kits, using the conditions recommended by the manufacturer (Qiagen Inc., Santa Clara, CA).

Template plasmid DNA clones are used for subsequent sequencing. ABI PRISM dRhodamine Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq™ DNA Polymerase, FS, is used for sequencing, (PE Applied Biosystems, Foster City, CA).

Example 3

DNA isolation and characterization

Plasmid DNA is isolated using the Promega Wizard™ Miniprep kit (Madison, WI), the Qiagen QIAwell Plasmid isolation kits (Chatsworth, GA) or Qiagen Plasmid Midi kit. These kits follow the same general procedure for plasmid DNA isolation. Briefly, cells are pelleted by centrifugation (5000 x g), plasmid DNA released with sequential NaOH/acid treatment, and cellular debris is removed by centrifugation (10000 x g). The supernatant (containing the plasmid DNA) is loaded onto a column containing a DNA-binding resin, the column is washed, and plasmid DNA eluted with TE. After screening for the colonies with the plasmid of interest, the *E. coli* cells are inoculated into 50-100 mls of LB plus appropriate antibiotic for overnight growth at 37°C in an air incubator while shaking. The

purified plasmid DNA is used for DNA sequencing, further restriction enzyme digestion, additional subcloning of DNA fragments and transfection or transduction into mammalian, *E. coli,* or other cells.

Sequence confirmation

Purified plasmid DNA is resuspended in dH_2O and quantitated by measuring the absorbance at 260/280 nm in a Bausch and Lomb Spectronic 601 UV spectrometer. DNA samples are sequenced using ABI PRISM™ DyeDeoxy™ terminator sequencing chemistry (Applied Biosystems Division of Perkin Elmer Corporation, Lincoln City, CA) kits (Part Number 401388 or 402078) according to the manufacturers suggested protocol usually modified by the addition of 5% DMSO to the sequencing mixture. Sequencing reactions are performed in a Model 480 DNA thermal cycler (Perkin Elmer Corporation, Norwalk, CT) following the recommended amplification conditions. Samples are purified to remove excess dye terminators with Centri-Sep™ spin columns (Princeton Separations, Adelphia, NJ) and lyophilized. Fluorescent dye labeled sequencing reactions are resuspended in deionized formamide, and sequenced on denaturing 4.75% polyacrylamide-8M urea gels using an ABI Model 373A automated DNA sequencer. Overlapping DNA sequence fragments are analyzed and assembled into master DNA contigs using Sequencher v3.0 DNA analysis software (Gene Codes Corporation, Ann Arbor, MI).

Expression of EGF motif repeat protein in mammalian cells

Mammalian Cell Transfection/Production of Conditioned Media

The BHK-21 cell line is obtained from the ATCC (Rockville, MD). The cells are cultured in Dulbecco's modified Eagle media (DMEM/high-glucose), supplemented to 2 mM (mM) L-glutamine and 10% fetal bovine serum (FBS). This formulation is designated BHK growth media. Selective media is BHK growth media supplemented with 453 units/mL hygromycin B (Calbiochem, San Diego, CA). The BHK-21 cell line is stably transfected with the HSV transactivating protein VP16, which transactivates the IE110 promoter found on the plasmid

pMON3359 (See Hippenmeyer *et al., Bio/Technology 11*: 1037-1041 (1993)). The VP16 protein drives expression of genes inserted behind the IE110 promoter. BHK-21 cells expressing the transactivating protein VP16 are designated BHK-VP16. The plasmid pMON1118 (See Highkin *et al., Poultry Sci. 70*:970-981 (1991)) expresses the hygromycin resistance gene from the SV40 promoter. A similar plasmid, available from ATCC, is pSV2-hph.

BHK-VP-16 cells are seeded into a 60 millimeter (mm) tissue culture dish at 3×10^5 cells per dish 24 hours prior to transfection. Cells are transfected for 16 hours in 3 mL of OPTIMEM™ (Gibco-BRL, Gaithersburg, MD) containing 10 μg of plasmid DNA containing the gene of interest, 3 μg hygromycin resistance plasmid, pMON1118, and 80 ug of Gibco-BRL LIPOFECTAMINE™ per dish. The media is subsequently aspirated and replaced with 3 mL of growth media. At 48 hours post-transfection, media from each dish is collected and assayed for activity (transient conditioned media). The cells are removed from the dish by trypsin-EDTA, diluted 1:10 and transferred to 100 mm tissue culture dishes containing 10 mL of selective media. After approximately 7 days in selective media, resistant cells grow into colonies several millimeters in diameter. The colonies are removed from the dish with filter paper (cut to approximately the same size as the colonies and soaked in trypsin/EDTA) and transferred to individual wells of a 24 wellplate containing 1 mL of selective media. After the clones are grown to confluence, the conditioned media is re-assayed, and positive clones are expanded into growth media.

Expression of EGF motif repeat protein in *E. coli*

E. coli strain MON105 or JM101 harboring the plasmid of interest are grown at 37°C in M9 plus casamino acids medium with shaking in an air incubator Model G25 from New Brunswick Scientific (Edison, New Jersey). Growth is monitored at OD_{600} until it reaches a value of 1.0 at which time Nalidixic acid (10 milligrams/mL) in 0.1 N NaOH is added to a final concentration of 50 μg/mL. The cultures are then shaken at 37°C for three to four additional hours. A high degree of aeration is maintained throughout culture period in order to achieve maximal

production of the desired gene product. The cells are examined under a light microscope for the presence of inclusion bodies (IB). One mL aliquots of the culture are removed for analysis of protein content by boiling the pelleted cells, treating them with reducing buffer and electrophoresis via SDS-PAGE (see Maniatis *et al., Molecular Cloning: A Laboratory Manual* (1982)). The culture is centrifuged (5000 x g) to pellet the cells.

Inclusion Body Preparation, Extraction, Refolding, Dialysis, DEAE Chromatography, and Characterization of the hematopoietic EGF motif

Isolation of Inclusion Bodies:

The cell pellet from a 330 mL *E. coli* culture is resuspended in 15 mL of sonication buffer (10 mM 2-amino-2-(hydroxymethyl) 1,3-propanediol hydrochloride (Tris-HC1), pH 8.0 + 1 mM ethylenediaminetetraacetic acid (EDTA). These rounds of sonication in sonication buffer followed by centrifugation are employed to disrupt the cells and wash the inclusion bodies (IB). The first round of sonication is a 3 minute burst followed by a 1 minute burst, and the final two rounds of sonication are for 1 minute each.

Extraction and refolding of proteins from inclusion body pellets:

Following the final centrifugation step, the IB pellet is resuspended in 10 mL of 50 mM Tris-HC1, pH 9.5, 8 M urea and 5 mM dithiothreitol (DTT) and stirred at room temperature for approximately 45 minutes to allow for denaturation of the expressed protein.

The extraction solution is transferred to a beaker containing 70 mL of 5 mM Tris-HC1, pH 9.5 and 2.3 M urea and gently stirred while exposed to air at 4°C for 18 to 48 hours to allow the proteins to refold. Refolding is monitored by analysis on a Vydac (Hesperia, Ca.) C18 reversed-phase-high pressure liquid chromatography (RP-HPLC) column (0.46x25 cm). A linear gradient of 40% to 65% acetonitrile, containing 0.15 trifluoracetic acid (TFA), is employed to monitor the refolding.

This gradient is developed over 30 minutes at a flow rate of 1.5 mL per minute. Denatured proteins generally elute later in the gradient than the refolded proteins.

Purification:

Following the refold, contaminating *E. coli* proteins are removed by acid precipitation. The pH of the refold solution is titrated to between pH 5.0 and pH 5.2 using 15% (v/v) acetic acid (HOAc). This solution is stirred at 4°C for 2 hours and then centrifuged for 20 minutes at 12,000 x g to pellet any insoluble protein.

The supernatant from the acid precipitation step is dialyzed using a Spectra/Por 3 membrane with a molecular weight cut off (MWCO) of 3,500 daltons. The dialysis is against 2 changes of 4 liters (a 50-fold excess) of 10 mM Tris-HCl, pH 8.0 for a total of 18 hours. Dialysis lowers the sample conductivity and removes urea prior to DEAE chromatography. The sample is then centrifuged (20 minutes at 12,000 x g) to pellet any insoluble protein following dialysis.

A Bio-Rad Bio-Scale DEAE2 column (7 x 52 mm) is used for ion exchange chromatography. The column is equilibrated in a buffer containing 10 mM Tris-HCl, pH 8.0, and a 0-to-500 mM sodium chloride (NaCl) gradient, in equilibration buffer, over 45 column volumes is used to elute the protein. A flow rate of 1.0 mL per minute is used throughout the run. Column fractions (2.0 mL per fraction) are collected across the gradient and analyzed by RP-HPLC on a Vydac (Hesperia, Ca.) C18 column (0.46 x 25 cm). A linear gradient of 40% to 65% acetonitrile, containing 0.1% trifluoroacetic acid (TFA), is employed. This gradient is developed over 30 minutes at a flow rate of 1.5 mL per minute. Pooled fractions are then dialyzed against 2 changes of 4 liters (50-to-500-fold excess) of 10 mM ammonium acetate (NH_4Ac), pH 4.0 for a total of 18 hours. Dialysis is performed using a Spectra/Por 3 membrane with a MWCO of 3,500 daltons. Finally, the sample is sterile filtered using a 0.22μm syringe filter (μStar LB syringe filter, Costar, Cambridge, Ma.), and stored at 4°C.

In some cases, the folded proteins can be affinity purified using affinity reagents such as mAbs or receptor subunits attached to a suitable matrix. Alternatively, (or in addition) purification can be accomplished using any of a

variety of chromatographic methods such as: ion exchange, gel filtration or hydrophobic chromatography or reversed phase HPLC.

These and other protein purification methods are described in detail in Deutgscher ed., *Methods in Enzymology*, Vol. 182 '*Guide to Protein Purification*' Academic Press, San Diego, CA (1990).

Protein Characterization:

The purified protein is analyzed by RP-HPLC, electrospray mass spectrometry, and SDS-PAGE. The protein quantitation is done by amino acid composition, RP-HPLC, and Bradford protein determination. In some cases tryptic peptide mapping is performed in conjunction with electrospray mass spectrometry to confirm the identity of the protein.

Additional details about recombinant DNA methods which may be used to create the variants, express them in bacteria, mammalian cells or insect cells, purification and refold of the desired proteins and assays for determining the bioactivity of the proteins may be found in WO 95/00646, WO 94/12639, WO 94/12638, WO 95/20976, WO 95/21197, WO 95/20977, and WO 95/21254 which are hereby incorporated by reference in their entirety.

Further details known to those skilled in the art may be found in Maniatis *et al.*, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, (1982); and in Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual*, 2nd edition, Cold Spring Harbor Laboratory, (1989).

Example 4

Generation of cDNA libraries

HS-5 cells were maintained in RPMI medium. 175cm^2 flasks were obtained, with approximately 1x10^7 HS-5 cells per flask. The cells were trypsinized and washed once in 1x PBS followed by resuspension in 12 ml of TRIzol reagent (Life Technologies).The expressed sequence tags were obtained by preparing RNA from

HS5 cells. Seventeen RNA was prepared using the standard protocol provided by the manufacturer (Life Technologies) with some modifications:

1. cells were lysed in TRIzol for 30 minutes at room temperature
2. RNA was precipitated and stored at -20C for nine days
3. A second TRIzol treatment of the RNA was used to remove residual contaminants.

The protocol yielded 1.5 mg of total RNA.

The total RNA was split into two identical aliquots for purification via a poly\A+ mRNA selection procedure. Mini-oligo dT cellulose spin columns (5prime→ 3prime Inc.) were used to isolate the polyA+ mRNA using the standard kit protocol specified by the manufacturer except poly A+ mRNA was twice selected by repeat passage on the oligo dT cellulose column. The yield from the protocol was 44 ug polyA+ mRNA.

Libraries were constructed from the mRNA using SuperScript™ Plasmid System for cDNA Synthesis and Plasmid Cloning (Life Technologies). The cDNA was synthesized from 3.5 ug polyA+. The library was prepared essentially according to the manufacturer's protocol with the following changes;

1. First strand synthesis for 60 minutes at 42C. Then, 1 ul SuperScript II was added and the reaction was incubated at 50C for 20 minutes.
2. T4 DNA ligase and NotI reactions were carried out using Boehringer Mannheim reagents.
3. cDNA was size-fractionated via 0.8% low melt SeaPlaque GTG agarose (FMC) gel in 1x TAE (Manitatis et al.) at 4C
4. Following the completion of the second strand synthesis, the SalI adapter ligation, the NotI digestion, and the size-fractionation, the cDNA was purified using GeneClean II (BIO 101) with two 10 ul elutions in water (65C).

The resulting cDNA was size fractionated and two separate pools were collected, 0.5-2.3k bp range and the 2.3k-7k bp range. The collection of cloned cDNA's was collectively referred to as library HS5. The library was transformed into E. coli and individual colonies were randomly selected for sequencing. Libraries designated

HS5R, HS5RODI, HS5RODI3, AND HS5RODI13 were prepared in a similar manner.

Subtractive library

Twice selected polyA+ mRNA was isolated from negative HS cell lines using the standard protocol for Poly (A) Pure kits (Cat # 1915) from Ambion (Austin, TX) . The subtracted libraries were constructed using Clontech's (Palo Alto, CA) PCR-Select cDNA Subtraction Kit (cat #K1804-1). HS-5 mRNA served as "tester" and HS-27 mRNA, a negative sister cell line, as "driver." After cDNA subtraction, the PCR amplified library's cDNA was ligated into the vector pCR2.1 from Invitrogen (Carlsbad, CA). The resulting library was designated HS527A and characterized by standard methods. The libraries designated LIB32, and LIB33 were prepared in a similar manner using the cell lines HS-27, and HS-23, respectively, as the negative cell line. The library HS5POOL was prepared in a similar manner using cell lines HS-21, HS-22, and HS-27 together as the negative cell lines. The library HS552122 was prepared in a similar manner using cell lines HS-21 and HS-22, together as the negative cell lines.

Example 5

Identification of HS-5 EST Candidates

HS5 Select

In the first method, all known growth factor sequences are extracted from a suitable database, such as the SwissProt database, using keyword searches and/or manual examination. The HS-5 ESTs are searched against the collection of known growth factors by TBLASTN, which searches a protein query against a DNA database by translating each database entry into all 6-reading frames. Alternatively, the HS-5 ESTs are searched against the collection of known growth factors by BLASTX, which translates a DNA query into putative peptides. There are five implementations of BLAST, three designed for nucleotide sequences queries (BLASTN, BLASTX, and TBLASTX) and two designed for protein sequence

queries (BLASTP and TBLASTN) (Coulson, *Trends in Biotechnology 12:* 76-80 (1994); Birren *et al., Genome Analysis 1:* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York 543-559 (1997)). BLASTN takes a nucleotide sequence (the query sequence) and its reverse complement and searches them against a nucleotide sequence database. BLASTX takes a nucleotide sequence, translates it in three forward reading frames and three reverse complement reading frames, and then compares the six translations against a protein sequence database. BLASTX is useful for sensitive analysis of preliminary (single-pass) sequence data and is tolerant of sequencing errors (Gish and States, *Nature Genetics 3:* 266-272 (1993), the entirety of which is herein incorporated by reference). BLASTN and BLASTX may be used in concert for analyzing EST data (Coulson, *Trends in Biotechnology 12:* 76-80 (1994); Birren *et al., Genome Analysis 1:* 543-559 (1997)).

The putative peptides are then searched against a suitable protein database. In this method, matches found with BLASTX values = 0.001 (probability) or a BLAST Score of = 90 are classified as hits.

In the second method, BLASTN searches are performed against the GenBank nonredundant nucleic acids database and BLASTX searches are performed against the GenBank nonredundant protein database. Since TBLASTN and BLASTX search the HS-5 database as putative protein translations, ESTs from untranslated regions of a mRNA (e.g., 5' UTR and 3' UTR) can be missed or misclassified due to nonsense peptide translations. The BLASTN search is designed to reduce the misclassification of ESTs due to nonsense peptide translations. TBLASTN searches can result in misclassification because of common features shared between growth factor and non-growth factor proteins. The BLASTX search is designed to reduce the misclassification of non-growth factor sequences as growth factor sequences because of common (and/or extraneous) features. Matches found with BLASTX or BLASTN with values = 0.0001 (probability) for nucleic acid or protein sequences known to be growth factors or associated with growth factor binding, signal transduction or proliferation are classified as hits.

Under the third method, the hidden Markov model (or HMMER) is used to detect more distant sequence similarities between the HS-5 candidates and known growth factor families (*see* Barrett *et al., Comput. Appl. Biosci. 13*: 191-199 (1997);

McClure *et al.*, *Ismb. 4*: 155-164 (1996)). Under this method, hidden Markov models of known growth factor families are used to search for growth factor-like sequences in the HS-5 EST translations. Hits with a score =5 are considered to be positive for growth factor sequence similarity. Where the HMMSW program (which is based on a hidden Markov model constructed to detect a specific gene family) is used, a HMMSW score = 10 is classified as a hit.

Because the putative HS-5 growth factor activity is present in HS-5 conditioned medium, the putative HS-5 growth factor is expected to be synthesized in the secretory pathway in the cell. Secreted proteins (which includes essentially all growth factors and growth factor receptors) share a leader sequence in the N-terminus called a signal peptide. Therefore, another method for identifying positive HS-5 ESTs encoding for the putative HS-5 growth factor is to identify HS-5 ESTs encoding for signal peptides. Two methods are employed for the identification of HS-5 ESTs encoding for signal peptides.

Novel Secreted HS5 (Indirect)

Under the indirect method, putative HS-5 growth factor ESTs identified using the database search methods are analyzed using SignalP, a neural network-based algorithm, to determine whether the ESTs encode for a signal peptide. Default parameters are used. If the scores for max. Y, max. S, and mean S are all above the cutoff (set by SignalP), a positive prediction is made.

Novel Secreted HS5 (Direct)

Under the direct method, sequences that did not receive any hits during the sequence database searches are analyzed using a two-step procedure. First, the GeneMark program is used to predict the coding regions of the EST sequences. The GeneMark program is a computer algorithm that identified characteristic features of nucleic acid sequences (*see* Isono *et al.*, *DNA Res. 1*: 263-269 (1994); McIninch *et al.*, *Ismb. 4*: 165-175 (1996); Lukashin *et al.*, *Nucleic Acids Res. 26*: 1107-1115 (1998)). Specifically, the GeneMark algorithm can be used to predict the exact boundaries of a gene. Coding regions are designated as those with a cutoff of P>0.5. Second, the coding regions identified by the GeneMark algorithm are

characterized by the SignalP program. If the scores for max. Y, max. S, and mean S are all above the cutoff (set by SignalP), a positive prediction is made.

Example 6

Nucleic acid sequences that encode proteins that are secreted from the HS5 cell line are identified by comparing to secreted protein sequences in the SwissProt database. Sequence comparisons between the SwissProt database and the HS-5 EST's are made with BLASTX, which translates the EST query in the six reading frames and compares the resultant six peptide sequences against the protein sequences in the boutique library. Matches found with BLASTX values equal or less than 0.001 (probability) or a BLAST Score of equal or greater than 90 are further classified by their annotation. SwissProt annotation contains a specific field labeled "Secreted" to designate a polypeptide that is known to be a secreted protein. The annotation of SwissProt matches is parsed to reveal whether the SwissProt match is a secreted protein. If an HS5 EST matched a SwissProt entry that is a secreted protein, the EST is classified as a hit

Example 7

Northern Blot Analysis of HS-5 ESTs

Northern analysis was performed on 12 ESTs of interest from the HS-5 cDNA library. To determine the full-length transcript size. Twice selected poly-A+ mRNA was prepared from HS-5 cells using the standard protocol for the Mini-Oligo (dT) Cellulose Spin Column Kit from 5 Prime – 3 Prime, Inc. Approximately one microgram HS-5 mRNA in denaturing buffer was electrophoresed per lane in a non-denaturing 1.25% agarose gel in 1X MOPS buffer (FMC BioProducts.) The gels also included RNA Ladder (Gibco BRL/Life Technologies) at five micrograms per well to provide size estimate. All samples were heat denatured before loading. The RNA gels were run in 1X MOPS buffer at approximately 3.5volts/cm gel length for two hours. Gels were washed for five minutes in deionized RNase free water. The portions of the gels containing the standards were removed for ethidium bromide

staining and subsequent size calibration. One lane from Gel#1 containing HS-5 mRNA was also removed for ethidium bromide staining to visualize the mRNA. The remaining portion of the gels was prepared for capillary transfer to nylon membranes. The gels were first soaked for 20 minutes in 0.05M sodium hydroxide. They were then soaked three times at 15 minutes each in 20X SSC buffer. Positively charged nylon was prepared by wetting first in deionized RNase free water followed by ten minutes in 20X SSC buffer. The capillary transfer was assembled in a standard manner (Molecular Cloning, A Laboratory Manual; Sambrook, et al.) using 20X SSC buffer. The transfer was allowed to proceed approximately 16 hours at room temperature. Upon completion of capillary transfer, the mRNA was irreversibly UV crosslinked to the nylon membrane. The nylon was cut into sections to provide one lane of HS-5 RNA per section. The nylon sections were hybridized individually with nonradioactive digoxigenin-labeled PCR products specific to the genes of interest. Probes were generated by standard protocol using purified plasmid DNA of each gene of interest as template, gene specific oligonucleotides, and PCR DIG Probe Synthesis Kit from Boehringer Mannheim. Membrane sections were incubated separately for one hour in DIG Easy Hyb Buffer (Boehringer Mannheim) at 50 degrees Celsius. Probes were heat denatured and utilized at two micrograms per milliliter, ten milliliters total volume DIG Easy Hybridization Buffer. Probe solution was added after the removal of prehybridization buffer. Incubation of probe and nylon was a minimum of 16 hours at 50 degrees Celsius with agitation. Sections were washed and developed per the DIG Wash Kit protocol, Boehringer Mannheim. Once autoradiographs were available, size correlations were drawn between the ethidium bromide stained RNA standards and the bands visible on the films. Figure 1 represents northern blot analysis of 12 EST clones and the asterisks in the figure represent the hybridization signal. The names of the clones are shown below each lane. The position of the size markers is shown by numbers and dashed lines. The size of the transcript was determined by extrapolation of a standard curve generated using the RNA marker.

Example 8

To identify a full-length cDNA of the EST clone LIB33-006-H9 (SEQ ID NO:1) a gene-specific primer pair was designed based on the sequence of EST clone LIB33-006-H9. L33-6-H9 sense primer sequence was AGGTACAAGGATCCCTGCGCTA SEQ ID NO:5, and L33-6-H9 antisense primer sequence was CACATCATAACATTGCAGGC SEQ ID NO:6. A probe was generated by amplifying a portion of the sequence using the primers L33-6-H9 sense and L33-6-H9 antisense and reagents from a Boehringer Mannheim PCR DIG Probe Synthesis kit to incorporate DIG-labeled dUTP into the PCR product. HS-5 cDNA clones were suspended in TB-amp^{100}, and 1 mL of the suspension was dispensed into each well of 12 X 96 deep well blocks (Qiagen cat # 26173) so there were ~1000 primary clones per well. The cultures were incubated at room temperature with shaking for 3 days, a fraction of each culture was mixed with glycerol and frozen, and plasmid DNA was isolated from the remainder of each culture. The plasmid DNA preps were robotically replicated onto positively charged nylon membrane (Boehringer Mannheim) in 96-well format. Approximately 1500 ng of DIG-labeled probe in approximately 100 mL DIG Easy Hyb (Boehringer Mannheim) was hybridized with the membrane at 42°C overnight. Washes and detection were performed according to DIG System protocols to identify wells that contained a clone with homology to the probe. A fraction of the freezer culture corresponding to each positive well was subcultured in TB-amp^{100} at 37°C with shaking overnight, and plasmid DNA was prepared from the overnight culture. One μg of the plasmid DNA was digested with restriction enzymes SalI and NotI to release inserts, the digestions were run on a 1% agarose gel, and transferred to positively charged nylon membrane (Boehringer Mannheim). Hybridization and detection were performed with the same probe solution and protocol used for the initial array to identify the size of the insert of the positive clone in each positive pool. One or more of the larger clones were isolated by titering out the frozen culture that was saved from each well, hybridizing against a colony lift of the titer, and subculturing a colony that aligned with a hybridization signal. The clone was sequenced by primer walking from the 5' and 3' ends. The resulting full-length clone was designated pMON37903 and contained the DNA sequence of SEQ ID NO:2, which encodes the deduced EGF motif repeat protein of SEQ IDNO:7.

SignalP analysis (Nielsen, H. et al., *Protein Engineering* **10**(1), 1997) predicted a probable signal peptide at the 5' end of the largest open reading frame of the sequence indicating that the clone was full-length.

All references and patents or applications cited herein are incorporated by reference in their entirety as if written herein.

Various other examples will be apparent to the person skilled in the art after reading the present disclosure without departing from the spirit and scope of the invention. It is intended that all such other examples be included within the scope of the appended claims.

WE CLAIM:

1. A substantially-purified nucleic acid molecule encoding an EGF motif repeat protein or fragment thereof, said nucleic acid molecule comprising at least one nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, substantial homologues thereof, and substantial fragments thereof.

2. The nucleic acid molecule according to claim 1, wherein said molecule shares at least 80% sequence identity, but for the degeneracy of the genetic code, with at least one nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and SEQ ID NO:4.

3. A substantially-purified nucleic acid molecule that encodes an EGF motif repeat protein or fragment thereof, said nucleic acid molecule having a nucleic acid sequence that hybridizes under stringent conditions or which would but for the degeneracy of the genetic code, with at least one nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and SEQ ID NO:4.

4. The nucleic acid of claim 1, 2, or 3 wherein the nucleic acid is selected from the group consisting of: SEQ ID NO:2, SEQ ID NO:3, and SEQ ID NO:4.

5. A EGF motif repeat protein having the amino acid sequence selected from the group consisting of SEQ ID NO:7, SEQ ID NO:8, a variant of SEQ ID NO:7, a variant of SEQ ID NO:8, a fragment of SEQ ID NO:7, and a fragment of SEQ ID NO:8.

6. A substantially purified antibody or fragment thereof, said antibody or fragment capable of specifically binding to the EGF motif repeat protein of claim 5.

Northern Analysis

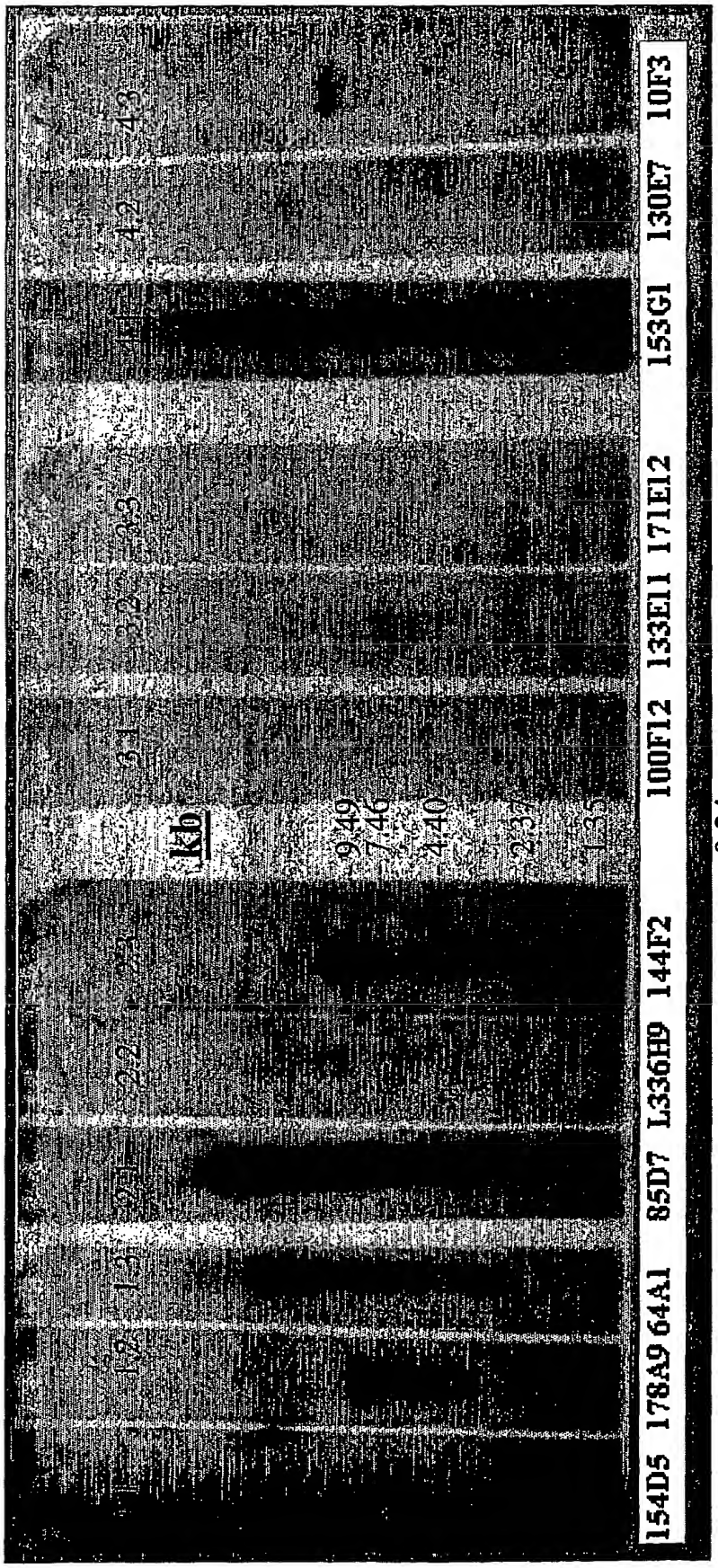


FIG 1

Figure 2a

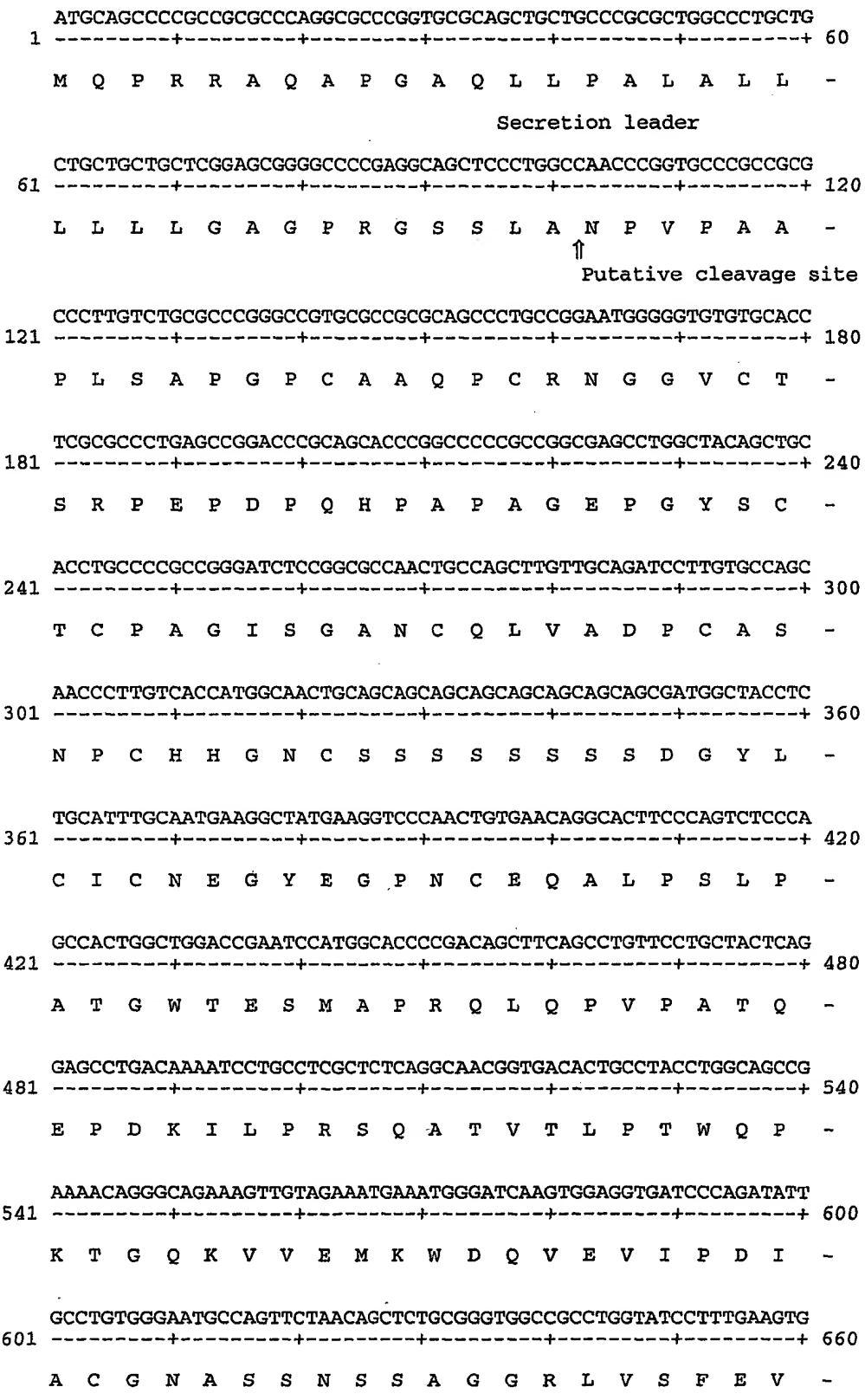

ATGCAGCCCCGCCGCGCCCAGGCGCCCGGTGCGCAGCTGCTGCCCGCGCTGGCCCTGCTG
1 60
M Q P R R A Q A P G A Q L L P A L A L L -
Secretion leader
CTGCTGCTGCTCGGAGCGGGGCCCCGAGGCAGCTCCCTGGCCAACCCGGTGCCCGCCGCG
61 120
L L L L G A G P R G S S L A N P V P A A -
Putative cleavage site
CCCTTGTCTGCGCCCGGGCCGTGCGCCGCGCAGCCCTGCCGGAATGGGGGTGTGTGCACC
121 180
P L S A P G P C A A Q P C R N G G V C T -
TCGCGCCCTGAGCCGGACCCGCAGCACCCGGCCCCCGCCGGCGAGCCTGGCTACAGCTGC
181 240
S R P E P D P Q H P A P A G E P G Y S C -
ACCTGCCCCGCCGGGATCTCCGGCGCCAACTGCCAGCTTGTTGCAGATCCTTGTGCCAGC
241 300
T C P A G I S G A N C Q L V A D P C A S -
AACCCTTGTCACCATGGCAACTGCAGCAGCAGCAGCAGCAGCAGCGATGGCTACCTC
301 360
N P C H H G N C S S S S S S S S D G Y L -
TGCATTTGCAATGAAGGCTATGAAGGTCCCAACTGTGAACAGGCACTTCCCAGTCTCCCA
361 420
C I C N E G Y E G P N C E Q A L P S L P -
GCCACTGGCTGGACCGAATCCATGGCACCCCGACAGCTTCAGCCTGTTCCTGCTACTCAG
421 480
A T G W T E S M A P R Q L Q P V P A T Q -
GAGCCTGACAAAATCCTGCCTCGCTCTCAGGCAACGGTGACACTGCCTACCTGGCAGCCG
481 540
E P D K I L P R S Q A T V T L P T W Q P -
AAAACAGGGCAGAAAGTTGTAGAAATGAAATGGGATCAAGTGGAGGTGATCCCAGATATT
541 600
K T G Q K V V E M K W D Q V E V I P D I -
GCCTGTGGGAATGCCAGTTCTAACAGCTCTGCGGGTGGCCGCCTGGTATCCTTTGAAGTG
601 660
A C G N A S S N S S A G G R L V S F E V -

Figure 2b

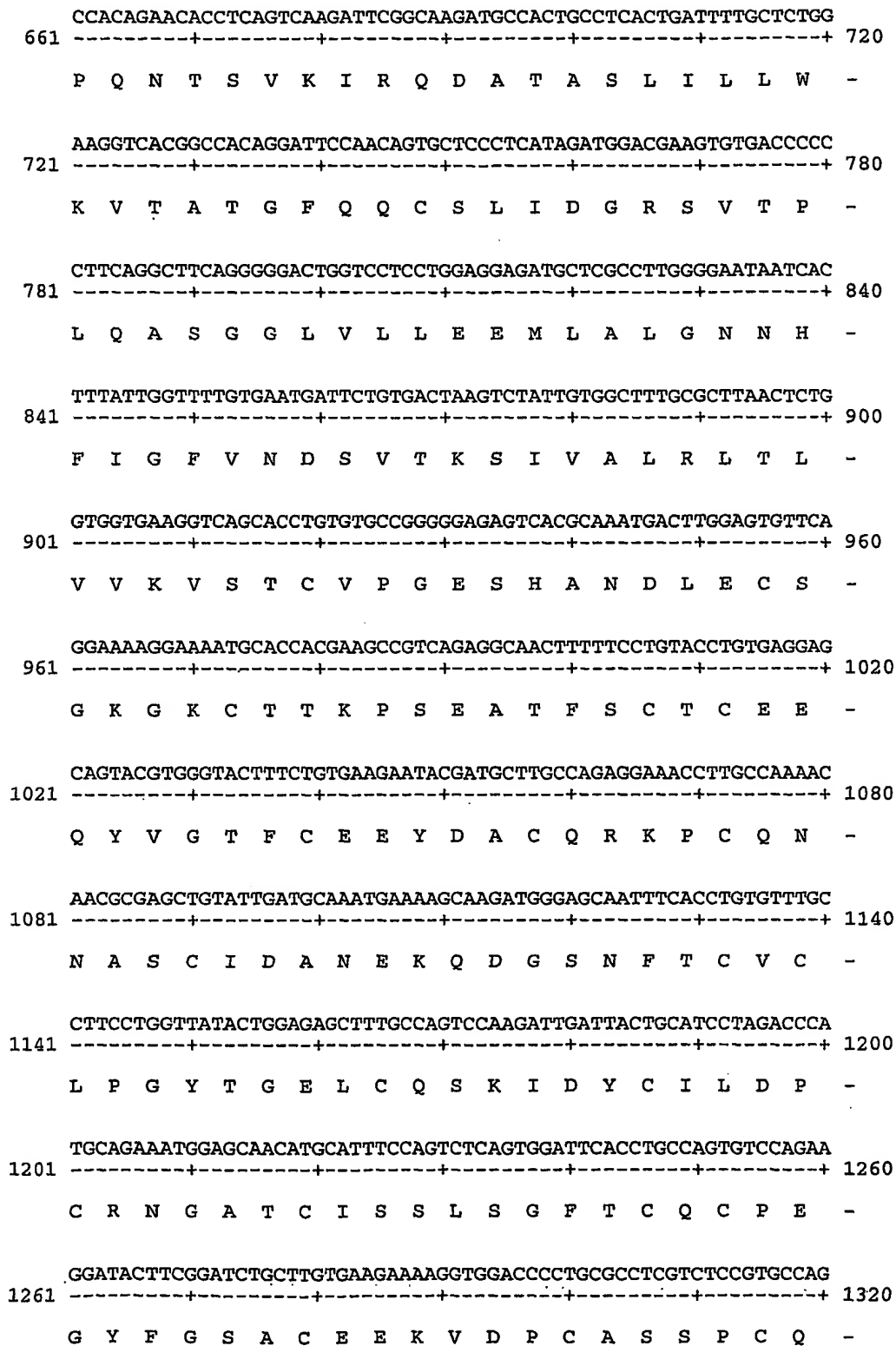

CCACAGAACACCTCAGTCAAGATTCGGCAAGATGCCACTGCCTCACTGATTTTGCTCTGG
661 ---------+---------+---------+---------+---------+---------+ 720
P Q N T S V K I R Q D A T A S L I L L W -
AAGGTCACGGCCACAGGATTCCAACAGTGCTCCCTCATAGATGGACGAAGTGTGACCCCC
721 ---------+---------+---------+---------+---------+---------+ 780
K V T A T G F Q Q C S L I D G R S V T P -
CTTCAGGCTTCAGGGGGACTGGTCCTCCTGGAGGAGATGCTCGCCTTGGGGAATAATCAC
781 ---------+---------+---------+---------+---------+---------+ 840
L Q A S G G L V L L E E M L A L G N N H -
TTTATTGGTTTTGTGAATGATTCTGTGACTAAGTCTATTGTGGCTTTGCGCTTAACTCTG
841 ---------+---------+---------+---------+---------+---------+ 900
F I G F V N D S V T K S I V A L R L T L -
GTGGTGAAGGTCAGCACCTGTGTGCCGGGGGAGAGTCACGCAAATGACTTGGAGTGTTCA
901 ---------+---------+---------+---------+---------+---------+ 960
V V K V S T C V P G E S H A N D L E C S -
GGAAAAGGAAAATGCACCACGAAGCCGTCAGAGGCAACTTTTTCCTGTACCTGTGAGGAG
961 ---------+---------+---------+---------+---------+---------+ 1020
G K G K C T T K P S E A T F S C T C E E -
CAGTACGTGGGTACTTTCTGTGAAGAATACGATGCTTGCCAGAGGAAACCTTGCCAAAAC
1021 ---------+---------+---------+---------+---------+---------+ 1080
Q Y V G T F C E E Y D A C Q R K P C Q N -
AACGCGAGCTGTATTGATGCAAATGAAAAGCAAGATGGGAGCAATTTCACCTGTGTTTGC
1081 ---------+---------+---------+---------+---------+---------+ 1140
N A S C I D A N E K Q D G S N F T C V C -
CTTCCTGGTTATACTGGAGAGCTTTGCCAGTCCAAGATTGATTACTGCATCCTAGACCCA
1141 ---------+---------+---------+---------+---------+---------+ 1200
L P G Y T G E L C Q S K I D Y C I L D P -
TGCAGAAATGGAGCAACATGCATTTCCAGTCTCAGTGGATTCACCTGCCAGTGTCCAGAA
1201 ---------+---------+---------+---------+---------+---------+ 1260
C R N G A T C I S S L S G F T C Q C P E -
GGATACTTCGGATCTGCTTGTGAAGAAAAGGTGGACCCCTGCGCCTCGTCTCCGTGCCAG
1261 ---------+---------+---------+---------+---------+---------+ 1320
G Y F G S A C E E K V D P C A S S P C Q -

Figure 2c

```
    AACAACGGCACCTGCTATGTGGACGGGGTACACTTTACCTGCAACTGCAGCCCGGGCTTC
1321 ---------+---------+---------+---------+---------+---------+ 1380

    N  N  G  T  C  Y  V  D  G  V  H  F  T  C  N  C  S  P  G  F  -

    ACAGGGCCGACCTGTGCCCAGCTTATTGACTTCTGTGCCCTCAGCCCCTGTGCTCATGGC
1381 ---------+---------+---------+---------+---------+---------+ 1440

    T  G  P  T  C  A  Q  L  I  D  F  C  A  L  S  P  C  A  H  G  -

    ACGTGCCGCAGCGTGGGCACCAGCTACAAATGCCTCTGTGATCCAGGTTACCATGGCCTC
1441 ---------+---------+---------+---------+---------+---------+ 1500

    T  C  R  S  V  G  T  S  Y  K  C  L  C  D  P  G  Y  H  G  L  -

    TACTGTGAGGAGGAATATAATGAGTGCCTCTCCGCTCCATGCCTGAATGCAGCCACCTGC
1501 ---------+---------+---------+---------+---------+---------+ 1560

    Y  C  E  E  E  Y  N  E  C  L  S  A  P  C  L  N  A  A  T  C  -

    AGGGACCTCGTTAATGGCTATGAGTGTGTGTGCCTGGCAGAATACAAAGGAACACACTGT
1561 ---------+---------+---------+---------+---------+---------+ 1620

    R  D  L  V  N  G  Y  E  C  V  C  L  A  E  Y  K  G  T  H  C  -

    GAATTGTACAAGGATCCCTGCGCTAACGTCAGCTGTCTGAACGGAGCCACCTGTGACAGC
1621 ---------+---------+---------+---------+---------+---------+ 1680

    E  L  Y  K  D  P  C  A  N  V  S  C  L  N  G  A  T  C  D  S  -

    GACGGCCTGAATGGCACGTGCATCTGTGCACCCGGGTTTACAGGTGAAGAGTGCGACATT
1681 ---------+---------+---------+---------+---------+---------+ 1740

    D  G  L  N  G  T  C  I  C  A  P  G  F  T  G  E  E  C  D  I  -

    GACATAAATGAATGTGACAGTAACCCCTGCCACCATGGTGGGAGCTGCCTGGACCAGCCC
1741 ---------+---------+---------+---------+---------+---------+ 1800

    D  I  N  E  C  D  S  N  P  C  H  H  G  G  S  C  L  D  Q  P  -

    AATGGTTATAACTGCCACTGCCCGCATGGTTGGGTGGGAGCAAACTGTGAGATCCACCTC
1801 ---------+---------+---------+---------+---------+---------+ 1860

    N  G  Y  N  C  H  C  P  H  G  W  V  G  A  N  C  E  I  H  L  -

    CAATGGAAGTCCGGGCACATGGCGGAGAGCCTCACCAACATGCCACGGCACTCCCTCTAC
1861 ---------+---------+---------+---------+---------+---------+ 1920

    Q  W  K  S  G  H  M  A  E  S  L  T  N  M  P  R  H  S  L  Y  -

    ATCATCATTGGAGCCCTCTGCGTGGCCTTCATCCTTATGCTGATCATCCTGATCGTGGGG
1921 ---------+---------+---------+---------+---------+---------+ 1980

    I  I  I  G  A  L  C  V  A  F  I  L  M  L  I  I  L  I  V  G  -
```

Figure 2d

```
    ATTTGCCGCATCAGCCGCATTGAATACCAGGGTTCTTCCAGGCCAGCCTATGAGGAGTTC
1981 ---------+---------+---------+---------+---------+---------+ 2040

    I  C  R  I  S  R  I  E  Y  Q  G  S  S  R  P  A  Y  E  E  F  -

    TACAACTGCCGCAGCATCGACAGCGAGTTCAGCAATGCCATTGCATCCATCCGGCATGCC
2041 ---------+---------+---------+---------+---------+---------+ 2100

    Y  N  C  R  S  I  D  S  E  F  S  N  A  I  A  S  I  R  H  A  -

    AGGTTTGGAAAGAAATCCCGGCCTGCAATGTATGATGTGAGCCCCATCGCCTATGAAGAT
2101 ---------+---------+---------+---------+---------+---------+ 2160

    R  F  G  K  K  S  R  P  A  M  Y  D  V  S  P  I  A  Y  E  D  -

    TACAGTCCTGATGACAAACCCTTGGTCACACTGATTAAAACTAAAGATTTGTAA   SEQ ID NO:3
2161 ---------+---------+---------+---------+---------+----      2214

    Y  S  P  D  D  K  P  L  V  T  L  I  K  T  K  D  L       SEQ ID NO:7
```

Figure 3a

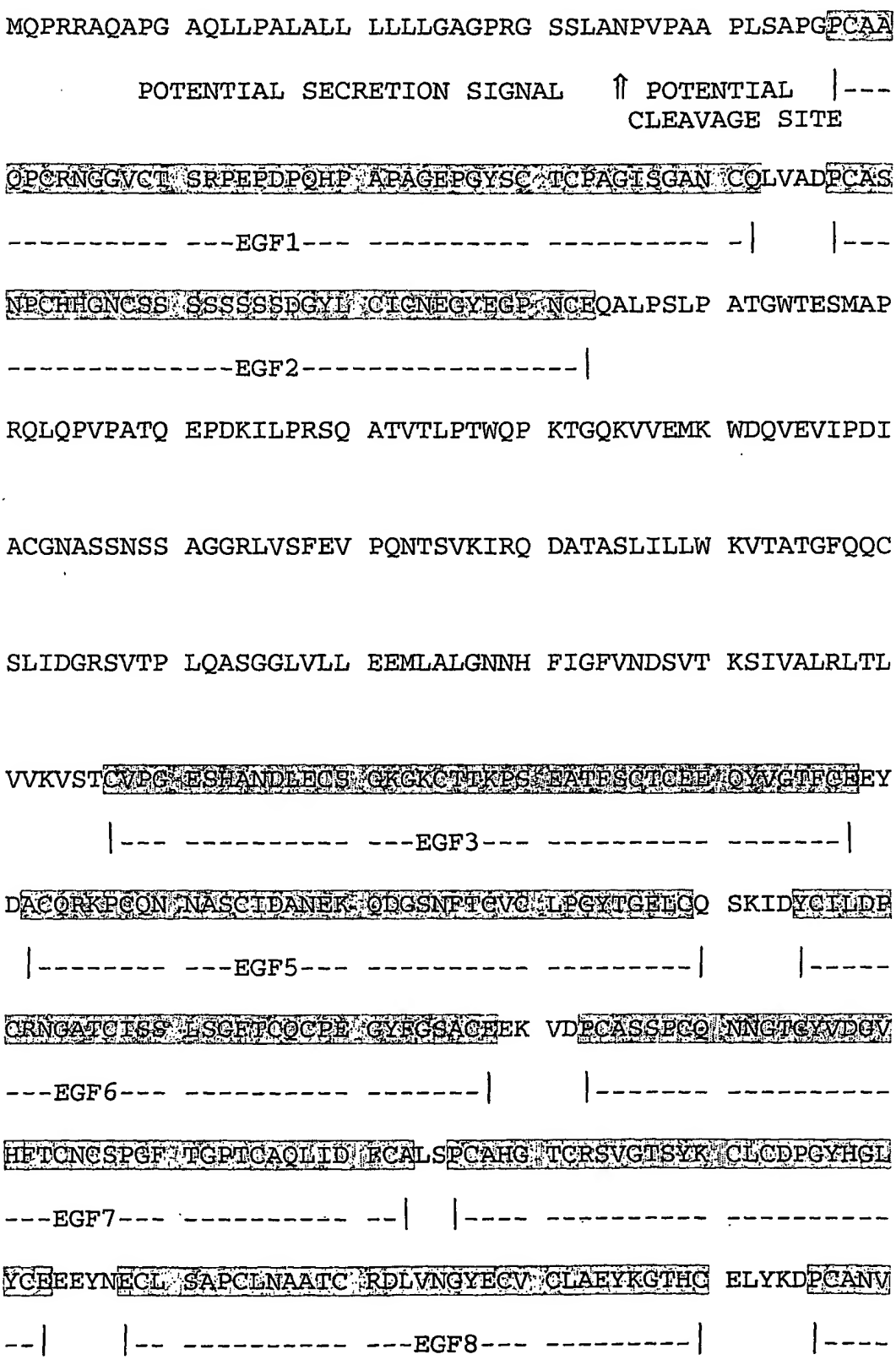
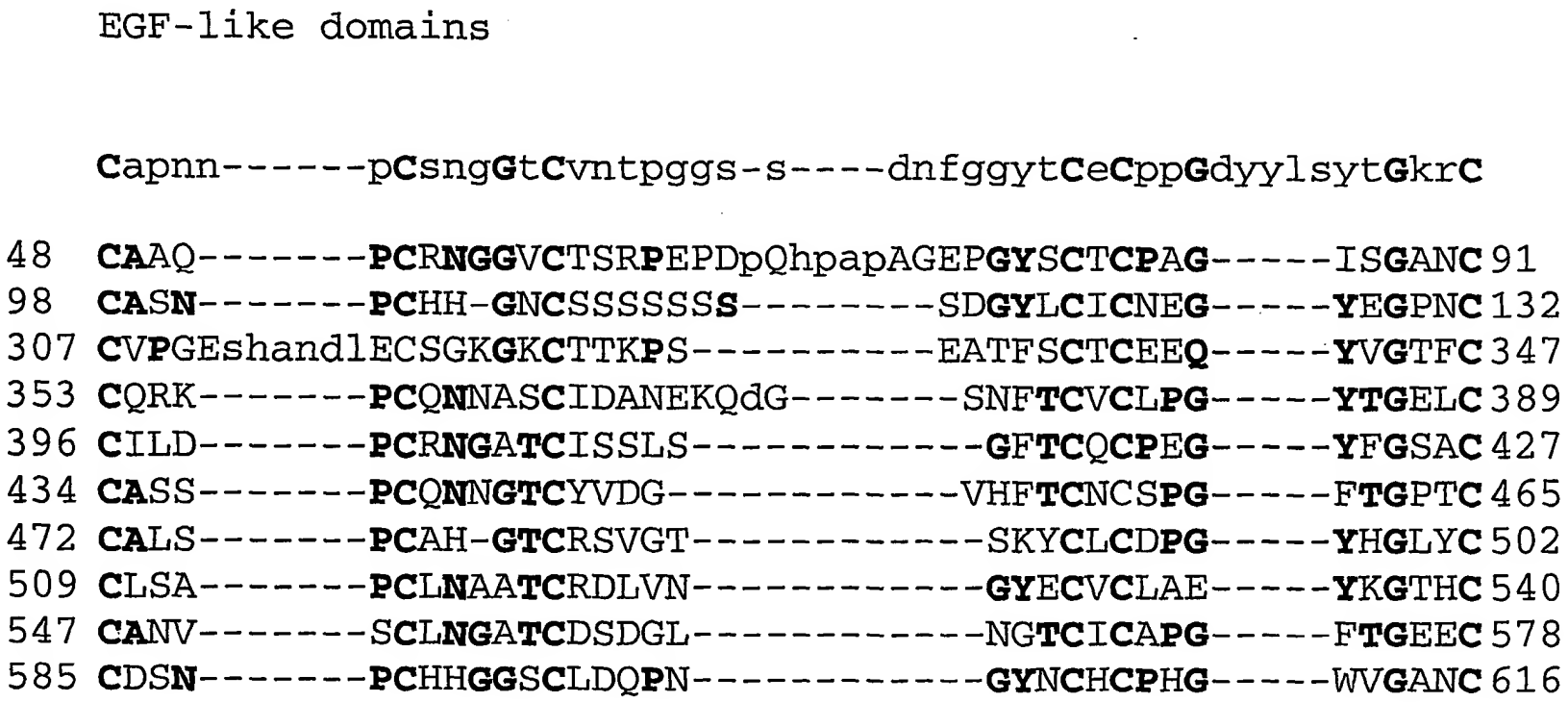

MQPRRAQAPG AQLLPALALL LLLLGAGPRG SSLANPVPAA PLSAPGPCAA 50
POTENTIAL SECRETION SIGNAL
POTENTIAL CLEAVAGE SITE
QPCRNGGVCT SRPEPDPQHP APAGEPGYSC TCPAGISGAN CQLVADPCAS 100
---EGF1---
NPCHHGNCSS SSSSSDGYL CICNEGYEGP NCEQALPSLP ATGWTESMAP 150
---EGF2---
RQLQPVPATQ EPDKILPRSQ ATVTLPTWQP KTGQKVVEMK WDQVEVIPDI 200
ACGNASSNSS AGGRLVSFEV PQNTSVKIRQ DATASLILLW KVTATGFQQC 250
SLIDGRSVTP LQASGGLVLL EEMLALGNNH FIGFVNDSVT KSIVALRLTL 300
VVKVSTCVPG ESHANDLECS GKGKCTTKPS EATFSCTCEE QYVGTFCEEY 350
---EGF3---
DACQRKPCQN NASCIDANEK QDGSNFTCVC LPGYTGELCQ SKIDYCILDE 400
---EGF5---
CRNGATCISS LSGFTCQCPE GYFGSACEEK VDPCASSPCQ NNGTCYVDGV 450
---EGF6---
HFTCNCSPGF TGPTCAQLID FCALSPCAHG TCRSVGTSYK CLCDPGYHGL 500
---EGF7---
YCEEEYNECL SAPCLNAATC RDLVNGYECV CLAEYKGTHC ELYKDPCANV 550
---EGF8---

Figure 3b

```
SCLNGATCDS DGLNGTCICA PGFTGEECDI DINECDSNPC HHGGSCLDQP 600
---EGF9--- ---------- --------|     |------ ---EGF10--
NGYNCHCPHG WVGANCEIHL QWKSGHMAES LTNMPRHSLY IIIGALCVAF 650
---------- ------|                       |- ---- TRANS
ILMLIILIVG ICRISRIEYQ GSSRPAYEEF YNCRSIDSEF SNAIASIRHA 700
MEMBRANE-- ---|
RFGKKSRPAM YDVSPIAYED YSPDDKPLVT LIKTKDL              737
```

SEQ ID NO:7

Figure 4


EGF-like domains
Capnn------pCsngGtCvntpggs-s----dnfggytCeCppGdyylsytGkrC
48 CAAQ-------PCRNGGVCTSRPEPDpQhpapAGEPGYSCTCPAG-----ISGANC 91
98 CASN-------PCHH-GNCSSSSSS--------SDGYLCICNEG-----YEGPNC 132
307 CVPGEshandlECSGKGKCTTKPS----------EATFSCTCEEQ-----YVGTFC 347
353 CQRK-------PCQNNASCIDANEKQdG-------SNFTCVCLPG-----YTGELC 389
396 CILD-------PCRNGATCISSLS----------GFTCQCPEG-----YFGSAC 427
434 CASS-------PCQNNGTCYVDG----------VHFTCNCSPG-----FTGPTC 465
472 CALS-------PCAH-GTCRSVGT----------SKYCLCDPG-----YHGLYC 502
509 CLSA-------PCLNAATCRDLVN----------GYECVCLAE-----YKGTHC 540
547 CANV-------SCLNGATCDSDGL----------NGTCICAPG-----FTGEEC 578
585 CDSN-------PCHHGGSCLDQPN----------GYNCHCPHG-----WVGANC 616

<110> Pharmacia
Doshi, Parul D.
Meier, Martin F.
Woods, Cynthia L.

<120> A NEW EGF MOTIF REPEAT PROTEIN OBTAINED FROM A CDNA LIBRARY FROM HS-5 STROMAL CELL LINE

<130> SO-3323/1/PCT

<140>

<141> 2001-08-10

<150> US 09/636,474

<151> 2000-08-10

<160> 8

<210> 1
<211> 394
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (1) . . . . (394)
<223> unsure at all n locations

<400> 1

acaaggatcc ctgcgctaac gtcagctgtc tgaacggagc cacctgtgac agcgacggcc 60

tgaatggcac gtgcatctgt gcacccgggt ttacaggtga agagtgcgac attgacataa 120

atgaatgtga cagtaacccc tgccaccatg gtgggagctg cctggaccag cccaatggtt 180

ataactgcca ctgcccgcat ggttgggtgg gagcaaactg tgagatccac ctccaatgga 240

agtccgggca catggcagag agcctcacca acatgccacg gcactccctc tacatcatca 300

ttggagccct ctgcgtggcc ttcatcctta tgctgatcat cctgatcgtg gggatttgcc 360

gcatcagccg cattgnataa cangggtctt ncag 394

<210> 2
<211> 3183
<212> DNA
<213> Homo sapiens

<400> 2

ggctcccgcg ccctcccggc cggccatgca gccccgccgc gcccaggcgc ccggtgcgca 60

gctgctgccc gcgctggccc tgctgctgct gctgctcgga gcggggcccc gaggcagctc 120

cctggccaac ccggtgcccg ccgcgccctt gtctgcgccc gggccgtgcg ccgcgcagcc 180

ctgccggaat gggggtgtgt gcacctcgcg ccctgagccg gacccgcagc acccggcccc 240

cgccggcgag cctggctaca gctgcacctg ccccgccggg atctccggcg ccaactgcca 300

gcttgttgca gatccttgtg ccagcaaccc ttgtcaccat ggcaactgca gcagcagcag 360

cagcagcagc agcgatggct acctctgcat ttgcaatgaa ggctatgaag gtcccaactg 420

tgaacaggca cttcccagtc tcccagccac tggctggacc gaatccatgg caccccgaca 480

gcttcagcct gttcctgcta ctcaggagcc tgacaaaatc ctgcctcgct ctcaggcaac 540

ggtgacactg cctacctggc agccgaaaac agggcagaaa gttgtagaaa tgaaatggga 600

tcaagtggag gtgatcccag atattgcctg tgggaatgcc agttctaaca gctctgcggg 660

tggccgcctg gtatcctttg aagtgccaca gaacacctca gtcaagattc ggcaagatgc 720

cactgcctca ctgattttgc tctggaaggt cacggccaca ggattccaac agtgctccct 780

catagatgga cgaagtgtga ccccccttca ggcttcaggg ggactggtcc tcctggagga 840

gatgctcgcc ttggggaata atcactttat tggttttgtg aatgattctg tgactaagtc 900

tattgtggct ttgcgcttaa ctctggtggt gaaggtcagc acctgtgtgc cgggggagag 960

tcacgcaaat gacttggagt gttcaggaaa aggaaaatgc accacgaagc cgtcagaggc 1020

aacttttcc tgtacctgtg aggagcagta cgtgggtact ttctgtgaag aatacgatgc 1080

ttgccagagg aaaccttgcc aaaacaacgc gagctgtatt gatgcaaatg aaaagcaaga 1140

tgggagcaat ttcacctgtg tttgccttcc tggttatact ggagagcttt gccagtccaa 1200

gattgattac tgcatcctag acccatgcag aaatggagca acatgcattt ccagtctcag 1260

tggattcacc tgccagtgtc cagaaggata cttcggatct gcttgtgaag aaaaggtgga 1320

cccctgcgcc tcgtctccgt gccagaacaa cggcacctgc tatgtggacg gggtacactt 1380

tacctgcaac tgcagcccgg gcttcacagg gccgacctgt gcccagctta ttgacttctg 1440

tgccctcagc ccctgtgctc atggcacgtg ccgcagcgtg ggcaccagct acaaatgcct 1500

ctgtgatcca ggttaccatg gcctctactg tgaggaggaa tataatgagt gcctctccgc 1560

tccatgcctg aatgcagcca cctgcaggga cctcgttaat ggctatgagt gtgtgtgcct 1620

ggcagaatac aaaggaacac actgtgaatt gtacaaggat ccctgcgcta acgtcagctg 1680

tctgaacgga gccacctgtg acagcgacgg cctgaatggc acgtgcatct gtgcacccgg 1740

gtttacaggt gaagagtgcg acattgacat aaatgaatgt gacagtaacc cctgccacca 1800

tggtgggagc tgcctggacc agcccaatgg ttataactgc cactgcccgc atggttgggt 1860

gggagcaaac tgtgagatcc acctccaatg gaagtccggg cacatggcgg agagcctcac 1920

caacatgcca cggcactccc tctacatcat cattggagcc ctctgcgtgg ccttcatcct 1980

tatgctgatc atcctgatcg tggggatttg ccgcatcagc cgcattgaat accagggttc 2040

ttccaggcca gcctatgagg agttctacaa ctgccgcagc atcgacagcg agttcagcaa 2100

tgccattgca tccatccggc atgccaggtt tggaaagaaa tcccggcctg caatgtatga 2160

tgtgagcccc atcgcctatg aagattacag tcctgatgac aaacccttgg tcacactgat 2220

taaaactaaa gatttgtaat ctttttttgg attatttttc aaaaagatga gatactacac 2280

tcatttaaat atttttaaga aaataaaaag cttaagaaat ttaaaatgct agctgctcaa 2340

gagttttcag tagaatattt aagaactaat tttctgcagc ttttagtttg gaaaaaatat 2400

tttaaaaaca aaatttgtga aacctataga cgatgtttta atgtaccttt agctctctaa 2460

actgtgtgct tctactagtg tgtgctcttt tcactgtaga cactatcacg agacccagat 2520

taatttctgt ggttgttaca gaataagtct aatcaaggag aagtttctgt ttgacgtttg 2580

agtgccggct ttctgagtag agttaggaaa accacgtaac gtagcatatg atgtataata 2640

gagtataccc gttacttaaa aagaagtctg aaatgttcgt tttgtggaaa agaaactagt 2700

taaatttact attcctaacc cgaatgaaat tagcctttgc cttattctgt gcatgggtaa 2760

gtaacttatt tctgcactgt tttgttgaac tttgtggaaa cattctttcg agtttgtttt 2820

tgtcattttc gtaacagtcg tcgaactagg cctcaaaaac atacgtaacg aaaaggccta 2880

gcgaggcaaa ttctgattga tttgaatcta tatttttctt taaaaagtca agggttctat 2940

attgtgagta aattaaattt acatttgagt tgtttgttgc taagaggtag taaatgtaag 3000

agagtactgg ttccttcagt agtgagtatt tctcatagtg cagctttatt tatctccagg 3060

atgtttttgt ggctgtattt gattgatatg tgcttcttct gattcttgct aatttccaac 3120

catattgaat aaatgtgatc aagtcaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 3180

aaa 3183

<210> 3
<211> 2214
<212> DNA
<213> Homo sapiens

<400> 3

atgcagcccc gccgcgccca ggcgcccggt gcgcagctgc tgcccgcgct ggccctgctg 60

ctgctgctgc tcggagcggg gccccgaggc agctccctgg ccaacccggt gcccgccgcg 120

cccttgtctg cgcccgggcc gtgcgccgcg cagccctgcc ggaatggggg tgtgtgcacc 180

tcgcgccctg agccggaccc gcagcacccg gcccccgccg gcgagcctgg ctacagctgc 240

acctgccccg ccgggatctc cggcgccaac tgccagcttg ttgcagatcc ttgtgccagc 300

aacccttgtc accatggcaa ctgcagcagc agcagcagca gcagcagcga tggctacctc 360

tgcatttgca atgaaggcta tgaaggtccc aactgtgaac aggcacttcc cagtctccca 420

gccactggct ggaccgaatc catggcaccc cgacagcttc agcctgttcc tgctactcag 480

gagcctgaca aaatcctgcc tcgctctcag gcaacggtga cactgcctac ctggcagccg 540

aaaacagggc agaaagttgt agaaatgaaa tgggatcaag tggaggtgat cccagatatt 600

gcctgtggga atgccagttc taacagctct gcgggtggcc gcctggtatc ctttgaagtg 660

ccacagaaca cctcagtcaa gattcggcaa gatgccactg cctcactgat tttgctctgg 720

aaggtcacgg ccacaggatt ccaacagtgc tccctcatag atggacgaag tgtgaccccc 780

cttcaggctt caggggggact ggtcctcctg gaggagatgc tcgccttggg gaataatcac 840

tttattggtt ttgtgaatga ttctgtgact aagtctattg tggctttgcg cttaactctg 900

gtggtgaagg tcagcacctg tgtgccgggg gagagtcacg caaatgactt ggagtgttca 960

ggaaaaggaa aatgcaccac gaagccgtca gaggcaactt tttcctgtac ctgtgaggag 1020

cagtacgtgg gtactttctg tgaagaatac gatgcttgcc agaggaaacc ttgccaaaac 1080

aacgcgagct gtattgatgc aaatgaaaag caagatggga gcaatttcac ctgtgtttgc 1140

cttcctggtt atactggaga gctttgccag tccaagattg attactgcat cctagaccca 1200

tgcagaaatg gagcaacatg catttccagt ctcagtggat tcacctgcca gtgtccagaa 1260

ggatacttcg gatctgcttg tgaagaaaag gtggacccct gcgcctcgtc tccgtgccag 1320

aacaacggca cctgctatgt ggacggggta cactttacct gcaactgcag cccgggcttc 1380

acagggccga cctgtgccca gcttattgac ttctgtgccc tcagcccctg tgctcatggc 1440

acgtgccgca gcgtgggcac cagctacaaa tgcctctgtg atccaggtta ccatggcctc 1500

tactgtgagg aggaatataa tgagtgcctc tccgctccat gcctgaatgc agccacctgc 1560

agggacctcg ttaatggcta tgagtgtgtg tgcctggcag aatacaaagg aacacactgt 1620

gaattgtaca aggatccctg cgctaacgtc agctgtctga acggagccac ctgtgacagc 1680

gacggcctga atggcacgtg catctgtgca cccgggttta caggtgaaga gtgcgacatt 1740

gacataaatg aatgtgacag taacccctgc caccatggtg ggagctgcct ggaccagccc 1800

aatggttata actgccactg cccgcatggt tgggtgggag caaactgtga gatccacctc 1860

caatggaagt ccgggcacat ggcggagagc ctcaccaaca tgccacggca ctccctctac 1920

atcatcattg gagccctctg cgtggccttc atccttatgc tgatcatcct gatcgtgggg 1980

atttgccgca tcagccgcat tgaataccag ggttcttcca ggccagccta tgaggagttc 2040

tacaactgcc gcagcatcga cagcgagttc agcaatgcca ttgcatccat ccggcatgcc 2100

aggtttggaa agaaatcccg gcctgcaatg tatgatgtga gccccatcgc ctatgaagat 2160

tacagtcctg atgacaaacc cttggtcaca ctgattaaaa ctaaagattt gtaa 2214

<210> 4
<211> 2102
<212> DNA
<213> Homo sapiens

<400> 4

aacccggtgc ccgccgcgcc cttgtctgcg cccgggccgt gcgccgcgca gccctgccgg 60

aatgggggtg tgtgcacctc gcgccctgag ccggacccgc agcacccggc ccccgccggc 120

gagcctggct acagctgcac ctgccccgcc gggatctccg gcgccaactg ccagcttgtt 180

gcagatcctt gtgccagcaa catggcaact gcagcagcag cagcagcagc agcagcgatg 240

gctacctctg catttgcaat gaaggctatg aaggtcccaa ctgtgaacag gcacttccca 300

gtctcccagc cactggctgg accgaatcca tggcaccccg acagcttcag cctgttcctg 360

ctactcagga gcctgacaaa atcctgcctc gctctcaggc aacggtgaca ctgcctacct 420

ggcagccgaa aacagggcag aaagttgtag aaatgaaatg ggatcaagtg gaggtgatcc 480

cagatattgc ctgtgggaat gccagttcta acagctctgc gggtggccgc ctggtatcct 540

ttgaagtgcc acagaacacc tcagtcaaga ttcggcaaga tgccactgcc tcactgattt 600

tgctctggaa ggtcacggcc acaggattcc aacagtgctc cctcatagat ggacgaagtg 660

tgacccccct tcaggcttca gggggactgg tcctcctgga ggagatgctc gccttgggga 720

ataatcactt tattggtttt gtgaatgatt ctgtgactaa gtctattgtg gctttgcgct 780

taactctggt ggtgaaggtc agcacctgtg tgccggggga gagtcacgca aatgacttgg 840

agtgttcagg aaaaggaaaa tgcaccacga agccgtcaga ggcaactttt tcctgtacct 900

gtgaggagca gtacgtgggt actttctgtg aagaatacga tgcttgccag aggaaacctt 960

gccaaaacaa cgcgagctgt attgatgcaa atgaaaagca agatgggagc aatttcacct 1020

gtgtttgcct tcctggttat actggagagc tttgccagtc caagattgat tactgcatcc 1080

tagacccatg cagaaatgga gcaacatgca tttccagtct cagtggattc acctgccagt 1140

gtccagaagg atacttcgga tctgcttgtg aagaaaaggt ggacccctgc gcctcgtctc 1200

cgtgccagaa caacggcacc tgctatgtgg acggggtaca ctttacctgc aactgcagcc 1260

cgggcttcac agggccgacc tgtgcccagc ttattgactt ctgtgccctc agcccctgtg 1320

ctcatggcac gtgccgcagc gtgggcacca gctacaaatg cctctgtgat ccaggttacc 1380

atggcctcta ctgtgaggag gaatataatg agtgcctctc cgctccatgc ctgaatgcag 1440

ccacctgcag ggacctcgtt aatggctatg agtgtgtgtg cctggcagaa tacaaaggaa 1500

cacactgtga attgtacaag gatccctgcg ctaacgtcag ctgtctgaac ggagccacct 1560

gtgacagcga cggcctgaat ggcacgtgca tctgtgcacc cgggtttaca ggtgaagagt 1620

gcgacattga cataaatgaa tgtgacagta acccctgcca ccatggtggg agctgcctgg 1680

accagcccaa tggttataac tgccactgcc cgcatggttg ggtgggagca aactgtgaga 1740

tccacctcca atggaagtcc gggcacatgg cggagagcct caccaacatg ccacggcact 1800

ccctctacat catcattgga gccctctgcg tggccttcat ccttatgctg atcatcctga 1860

tcgtggggat ttgccgcatc agccgcattg aataccaggg ttcttccagg ccagcctatg 1920

aggagttcta caactgccgc agcatcgaca gcgagttcag caatgccatt gcatccatcc 1980

ggcatgccag gtttggaaag aaatcccggc ctgcaatgta tgatgtgagc cccatcgcct 2040

atgaagatta cagtcctgat gacaaaccct tggtcacact gattaaaact aaagatttgt 2100

aa 2102

<210> 5
<211> 22
<212> DNA
<213> Homo sapiens

<400> 5

aggtacaagg atccctgcgc ta 22

<210> 6
<211> 20
<212> DNA
<213> Homo sapiens

<400> 6

cacatcataa cattgcaggc 20

<210> 7
<211> 737
<212> Prt

<213> Homo sapiens

<400> 7

Met Gln Pro Arg Arg Ala Gln Ala Pro Gly Ala Gln Leu Leu Pro
1 5 10 15

Ala Leu Ala Leu Leu Leu Leu Leu Leu Gly Ala Gly Pro Arg Gly
20 25 30

Ser Ser Leu Ala Asn Pro Val Pro Ala Ala Pro Leu Ser Ala Pro
35 40 45

Gly Pro Cys Ala Ala Gln Pro Cys Arg Asn Gly Gly Val Cys Thr
50 55 60

Ser Arg Pro Glu Pro Asp Pro Gln His Pro Ala Pro Ala Gly Glu
65 70 75

Pro Gly Tyr Ser Cys Thr Cys Pro Ala Gly Ile Ser Gly Ala Asn
80 85 90

Cys Gln Leu Val Ala Asp Pro Cys Ala Ser Asn Pro Cys His His
95 100 105

Gly Asn Cys Ser Ser Ser Ser Ser Ser Ser Ser Asp Gly Tyr Leu
110 115 120

Cys Ile Cys Asn Glu Gly Tyr Glu Gly Pro Asn Cys Glu Gln Ala
125 130 135

Leu Pro Ser Leu Pro Ala Thr Gly Trp Thr Glu Ser Met Ala Pro
140 145 150

Arg Gln Leu Gln Pro Val Pro Ala Thr Gln Glu Pro Asp Lys Ile
155 160 165

Leu Pro Arg Ser Gln Ala Thr Val Thr Leu Pro Thr Trp Gln Pro
170 175 180

Lys Thr Gly Gln Lys Val Val Glu Met Lys Trp Asp Gln Val Glu
185 190 195

Val Ile Pro Asp Ile Ala Cys Gly Asn Ala Ser Ser Asn Ser Ser
200 205 210

Ala Gly Gly Arg Leu Val Ser Phe Glu Val Pro Gln Asn Thr Ser
215 220 225

Val Lys Ile Arg Gln Asp Ala Thr Ala Ser Leu Ile Leu Leu Trp
230 235 240

Lys Val Thr Ala Thr Gly Phe Gln Gln Cys Ser Leu Ile Asp Gly
245 250 255

Arg Ser Val Thr Pro Leu Gln Ala Ser Gly Gly Leu Val Leu Leu
260 265 270

Glu Glu Met Leu Ala Leu Gly Asn Asn His Phe Ile Gly Phe Val

275 280 285

Asn Asp Ser Val Thr Lys Ser Ile Val Ala Leu Arg Leu Thr Leu
290 295 300

Val Val Lys Val Ser Thr Cys Val Pro Gly Glu Ser His Ala Asn
305 310 315

Asp Leu Glu Cys Ser Gly Lys Gly Lys Cys Thr Thr Lys Pro Ser
320 325 330

Glu Ala Thr Phe Ser Cys Thr Cys Glu Glu Gln Tyr Val Gly Thr
335 340 345

Phe Cys Glu Glu Tyr Asp Ala Cys Gln Arg Lys Pro Cys Gln Asn
350 355 360

Asn Ala Ser Cys Ile Asp Ala Asn Glu Lys Gln Asp Gly Ser Asn
365 370 375

Phe Thr Cys Val Cys Leu Pro Gly Tyr Thr Gly Glu Leu Cys Gln
380 385 390

Ser Lys Ile Asp Tyr Cys Ile Leu Asp Pro Cys Arg Asn Gly Ala
395 400 405

Thr Cys Ile Ser Ser Leu Ser Gly Phe Thr Cys Gln Cys Pro Glu
410 415 420

Gly Tyr Phe Gly Ser Ala Cys Glu Glu Lys Val Asp Pro Cys Ala
425 430 435

Ser Ser Pro Cys Gln Asn Asn Gly Thr Cys Tyr Val Asp Gly Val
440 445 450

His Phe Thr Cys Asn Cys Ser Pro Gly Phe Thr Gly Pro Thr Cys
455 460 465

Ala Gln Leu Ile Asp Phe Cys Ala Leu Ser Pro Cys Ala His Gly
470 475 480

Thr Cys Arg Ser Val Gly Thr Ser Tyr Lys Cys Leu Cys Asp Pro
485 490 495

Gly Tyr His Gly Leu Tyr Cys Glu Glu Glu Tyr Asn Glu Cys Leu
500 505 510

Ser Ala Pro Cys Leu Asn Ala Ala Thr Cys Arg Asp Leu Val Asn
515 520 525

Gly Tyr Glu Cys Val Cys Leu Ala Glu Tyr Lys Gly Thr His Cys
530 535 540

Glu Leu Tyr Lys Asp Pro Cys Ala Asn Val Ser Cys Leu Asn Gly
545 550 555

Ala Thr Cys Asp Ser Asp Gly Leu Asn Gly Thr Cys Ile Cys Ala
560 565 570

```
Pro Gly Phe Thr Gly Glu Glu Cys Asp Ile Asp Ile Asn Glu Cys
                575                 580                 585

Asp Ser Asn Pro Cys His His Gly Gly Ser Cys Leu Asp Gln Pro
                590                 595                 600

Asn Gly Tyr Asn Cys His Cys Pro His Gly Trp Val Gly Ala Asn
                605                 610                 615

Cys Glu Ile His Leu Gln Trp Lys Ser Gly His Met Ala Glu Ser
                620                 625                 630

Leu Thr Asn Met Pro Arg His Ser Leu Tyr Ile Ile Ile Gly Ala
                635                 640                 645

Leu Cys Val Ala Phe Ile Leu Met Leu Ile Ile Leu Ile Val Gly
                650                 655                 660

Ile Cys Arg Ile Ser Arg Ile Glu Tyr Gln Gly Ser Ser Arg Pro
                665                 670                 675

Ala Tyr Glu Glu Phe Tyr Asn Cys Arg Ser Ile Asp Ser Glu Phe
                680                 685                 690

Ser Asn Ala Ile Ala Ser Ile Arg His Ala Arg Phe Gly Lys Lys
                695                 700                 705

Ser Arg Pro Ala Met Tyr Asp Val Ser Pro Ile Ala Tyr Glu Asp
                710                 715                 720

Tyr Ser Pro Asp Asp Lys Pro Leu Val Thr Leu Ile Lys Thr Lys
                725                 730                 735

Asp Leu
    737

<210>       8
<211>       703
<212>       Prt
<213>       Homo sapiens

<400>       8

Asn Pro Val Pro Ala Ala Pro Leu Ser Ala Pro Gly Pro Cys Ala
1               5                   10                  15

Ala Gln Pro Cys Arg Asn Gly Gly Val Cys Thr Ser Arg Pro Glu
                20                  25                  30

Pro Asp Pro Gln His Pro Ala Pro Ala Gly Glu Pro Gly Tyr Ser
                35                  40                  45

Cys Thr Cys Pro Ala Gly Ile Ser Gly Ala Asn Cys Gln Leu Val
                50                  55                  60

Ala Asp Pro Cys Ala Ser Asn Pro Cys His His Gly Asn Cys Ser
                65                  70                  75

Ser Ser Ser Ser Ser Ser Ser Asp Gly Tyr Leu Cys Ile Cys Asn
```

```
                80                  85                  90

Glu Gly Tyr Glu Gly Pro Asn Cys Glu Gln Ala Leu Pro Ser Leu
                95                 100                 105

Pro Ala Thr Gly Trp Thr Glu Ser Met Ala Pro Arg Gln Leu Gln
               110                 115                 120

Pro Val Pro Ala Thr Gln Glu Pro Asp Lys Ile Leu Pro Arg Ser
               125                 130                 135

Gln Ala Thr Val Thr Leu Pro Thr Trp Gln Pro Lys Thr Gly Gln
               140                 145                 150

Lys Val Val Glu Met Lys Trp Asp Gln Val Glu Val Ile Pro Asp
               155                 160                 165

Ile Ala Cys Gly Asn Ala Ser Ser Asn Ser Ser Ala Gly Gly Arg
               170                 175                 180

Leu Val Ser Phe Glu Val Pro Gln Asn Thr Ser Val Lys Ile Arg
               185                 190                 195

Gln Asp Ala Thr Ala Ser Leu Ile Leu Leu Trp Lys Val Thr Ala
               200                 205                 210

Thr Gly Phe Gln Gln Cys Ser Leu Ile Asp Gly Arg Ser Val Thr
               215                 220                 225

Pro Leu Gln Ala Ser Gly Gly Leu Val Leu Leu Glu Glu Met Leu
               230                 235                 240

Ala Leu Gly Asn Asn His Phe Ile Gly Phe Val Asn Asp Ser Val
               245                 250                 255

Thr Lys Ser Ile Val Ala Leu Arg Leu Thr Leu Val Val Lys Val
               260                 265                 270

Ser Thr Cys Val Pro Gly Glu Ser His Ala Asn Asp Leu Glu Cys
               275                 280                 285

Ser Gly Lys Gly Lys Cys Thr Thr Lys Pro Ser Glu Ala Thr Phe
               290                 295                 300

Ser Cys Thr Cys Glu Glu Gln Tyr Val Gly Thr Phe Cys Glu Glu
               305                 310                 315

Tyr Asp Ala Cys Gln Arg Lys Pro Cys Gln Asn Asn Ala Ser Cys
               320                 325                 330

Ile Asp Ala Asn Glu Lys Gln Asp Gly Ser Asn Phe Thr Cys Val
               335                 340                 345

Cys Leu Pro Gly Tyr Thr Gly Glu Leu Cys Gln Ser Lys Ile Asp
               350                 355                 360

Tyr Cys Ile Leu Asp Pro Cys Arg Asn Gly Ala Thr Cys Ile Ser
               365                 370                 375
```

Ser Leu Ser Gly Phe Thr Cys Gln Cys Pro Glu Gly Tyr Phe Gly
380 385 390

Ser Ala Cys Glu Glu Lys Val Asp Pro Cys Ala Ser Ser Pro Cys
395 400 405

Gln Asn Asn Gly Thr Cys Tyr Val Asp Gly Val His Phe Thr Cys
410 415 420

Asn Cys Ser Pro Gly Phe Thr Gly Pro Thr Cys Ala Gln Leu Ile
425 430 435

Asp Phe Cys Ala Leu Ser Pro Cys Ala His Gly Thr Cys Arg Ser
440 445 450

Val Gly Thr Ser Tyr Lys Cys Leu Cys Asp Pro Gly Tyr His Gly
455 460 465

Leu Tyr Cys Glu Glu Glu Tyr Asn Glu Cys Leu Ser Ala Pro Cys
470 475 480

Leu Asn Ala Ala Thr Cys Arg Asp Leu Val Asn Gly Tyr Glu Cys
485 490 495

Val Cys Leu Ala Glu Tyr Lys Gly Thr His Cys Glu Leu Tyr Lys
500 505 510

Asp Pro Cys Ala Asn Val Ser Cys Leu Asn Gly Ala Thr Cys Asp
515 520 525

Ser Asp Gly Leu Asn Gly Thr Cys Ile Cys Ala Pro Gly Phe Thr
530 535 540

Gly Glu Glu Cys Asp Ile Asp Ile Asn Glu Cys Asp Ser Asn Pro
545 550 555

Cys His His Gly Gly Ser Cys Leu Asp Gln Pro Asn Gly Tyr Asn
560 565 570

Cys His Cys Pro His Gly Trp Val Gly Ala Asn Cys Glu Ile His
575 580 585

Leu Gln Trp Lys Ser Gly His Met Ala Glu Ser Leu Thr Asn Met
590 595 600

Pro Arg His Ser Leu Tyr Ile Ile Ile Gly Ala Leu Cys Val Ala
605 610 615

Phe Ile Leu Met Leu Ile Ile Leu Ile Val Gly Ile Cys Arg Ile
620 625 630

Ser Arg Ile Glu Tyr Gln Gly Ser Ser Arg Pro Ala Tyr Glu Glu
635 640 645

Phe Tyr Asn Cys Arg Ser Ile Asp Ser Glu Phe Ser Asn Ala Ile
650 655 660

Ala Ser Ile Arg His Ala Arg Phe Gly Lys Lys Ser Arg Pro Ala
665 670 675

Met Tyr Asp Val Ser Pro Ile Ala Tyr Glu Asp Tyr Ser Pro Asp
680 685 690

Asp Lys Pro Leu Val Thr Leu Ile Lys Thr Lys Asp Leu
695 700 703